Childhood confidence, schooling, and the labor market: Evidence from the PSID

Lucy Page and Hannah Ruebeck*

MIT

March 2022

Abstract

We link over- and under-confidence in math at ages 8-11 to education and employment outcomes 22 years later among the children of PSID households. About twenty percent of children have markedly biased beliefs about their math ability, and beliefs are strongly gendered. Childhood over- and under-confidence predict substantial gaps in later test scores, high school and college graduation, majoring or working in STEM, earnings, and unemployment, all estimated with extensive controls for ability. Across all metrics, higher confidence predicts better outcomes. These biased beliefs persist into adulthood and could continue to affect outcomes as respondents age.

*Page: PhD candidate, MIT Economics, Cambridge, MA 02142, lucypage@mit.edu. Ruebeck: PhD candidate, MIT Economics, Cambridge, MA 02142, hruebeck@mit.edu (corresponding author). We are grateful to David Autor, Esther Duflo, Amy Finkelstein, Kartini Shastry, Rohini Pande, and Frank Schilbach for thoughtful and helpful comments. Both authors are supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. This project is also supported by the George and Obie Shultz Fund at MIT.

The majority of data used in this paper is publicly available on the website of the Panel Study of Income Dynamics: https://psidonline.isr.umich.edu/. Several control and outcome variables used in the analysis rely on the restricted PSID dataset, which can be obtained via the process described at https://simba.isr.umich.edu/restricted/ProcessReq.aspx.
March 1, 2022

To Whom It May Concern:

The manuscript “Childhood confidence, schooling, and the labor market: Evidence from the PSID” was written with financial support from the George and Obie Shultz Fund at MIT. Lucy Page was supported by a National Science Foundation Graduate Research Fellowship throughout the duration of this project. IRB approval for the project was obtained from the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects (Protocol # 1910000016). No other party had the right to review the paper prior to its circulation. Lucy Page declares that she has no relevant or material financial interests that relate to the research described in this paper.

Sincerely,

[Signature]
Lucy Page
March 1, 2022

To Whom It May Concern:

The manuscript “Childhood confidence, schooling, and the labor market: Evidence from the PSID” was written with financial support from the George and Obie Shultz Fund at MIT. Hannah Ruebeck was supported by a National Science Foundation Graduate Research Fellowship throughout the duration of this project. IRB approval for the project was obtained from the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects (Protocol # 1910000016). No other party had the right to review the paper prior to its circulation. Hannah Ruebeck declares that she has no relevant or material financial interests that relate to the research described in this paper.

Sincerely,

Hannah Ruebeck
1 Introduction

Long-standing research in psychology finds that people have biased beliefs about their abilities in a range of domains.¹ Prior research has focused on “optimism bias,” or over-confidence about one’s performance, belief accuracy, or future outcomes (Moore and Healy, 2008; Sharot et al., 2011; Taylor and Brown, 1988). In contrast, psychologists also document “imposter syndrome,” a form of systematic under-confidence in which people attribute their successes to luck or effort rather than skill (Langford and Clance, 1993; Sakulku, 2011). Recent lab-based work in behavioral economics has sought to microfound this empirical evidence of biased beliefs by documenting that people systematically under-weight or over-weight signals about the truth, especially in ego-relevant domains like intelligence and beauty (see Benjamin (2019) for a review).

Do these confidence gaps matter for economic decision-making in the real world? There are key reasons to expect that they might. For example, if adolescents or young adults perceive ability and educational investment to be complements, under-confident students might exert less effort in school or end their education earlier (Bénabou and Tirole, 2002). Later, under-confident adults may be less likely to complete costly and uncertain job applications, or may select away from jobs with higher returns to performance (Dohmen and Falk, 2011).² Individuals’ beliefs about their own ability could also affect their outcomes by shaping how others perceive them. If parents or teachers perceive more confident students as higher-ability and expect the returns of education to increase with ability, they may invest more in more confident children (Papageorge et al., 2018; Dizon-Ross, 2019). More confident applicants may appear more capable during job interviews, improving their earnings and employment prospects (Schwardmann and van der Weele, 2019).

As yet, there is limited evidence for how confidence affects economic outcomes in realistic settings and over the long term. In addition to the lab-based work on the short-term implications of

---

¹We refer throughout the paper to ability and beliefs about ability, but we do not mean to imply that ability (or beliefs) are innate or fixed. Rather, we are referring to someone’s ability or perceived ability to perform well in a certain domain or task at a particular time.

²Psychological theories of motivation, including Bandura’s (1986) Social Cognitive Theory or Expectancy-Value Theory (see Wigfield and Eccles (2000)) also emphasize that individuals increase their effort in domains in which they feel competent.
confidence gaps cited above (Dohmen and Falk, 2011; Schwardmann and van der Weele, 2019), a small parallel literature in economics and sociology examines longer-term outcomes and finds that those with higher self-esteem get more education, are more likely to be employed, and earn higher wages (Murnane et al. 2001; Waddell 2006; Drago 2011; de Araujo and Lagos, 2013). However, this literature has struggled to demonstrate that these associations are not driven by omitted variables like unobserved ability. These papers typically control for IQ in an attempt to account for cognitive ability, but it is not feasible to control for subjects’ “ability” across all domains that affect generalized self-esteem.

In this paper, we address the limitations of both prior literatures by examining the real-world and long-term implications of a dimension of confidence in which we can observe and control for demonstrated ability: childhood over- and under-confidence in math.\(^3\) We use unique data from the Panel Study of Income Dynamics (PSID) to identify biased beliefs in math in a sample of 2,985 children in core PSID households; we then relate their childhood over- and under-confidence to educational and employment outcomes up to 22 years later, controlling for test scores, general confidence, and other key confounders.

The PSID is an ideal setting in which to examine long-term links with childhood confidence. Our sample is based on child-focused PSID supplements that measure both children’s performance on a standardized math test and their own assessments of how “good” they are at math. We combine these measures to identify over-confident children as those who scored poorly on the math assessment and yet said they were good at math, and to identify under-confident children as those who scored well on the assessment but said they were bad at math. The structure of the PSID also allows us to observe much of respondents’ young adulthood: the child supplements and core survey followed our sample from 1997 through 2019, so we observe our oldest respondents from age 12 into their mid thirties.

Biased beliefs about math ability are prevalent in our sample: 5-20 percent of students are

---

\(^3\) We also construct parallel measures of reading over- and under-confidence, and we report all of the following analysis for reading confidence in Appendix Tables A2-A5. We discuss our focus on math confidence in Section 2.
markedly over-confident and 7-16 percent of are markedly under-confident (using several definitions of biased beliefs, described in more detail below). Over- and under-confidence in math are also highly gendered: girls are 2.3 percentage points (pp) (17 percent) more likely to be under-confident and 2.7 pp (27 percent) less likely to be over-confident in math than boys. In contrast, girls are 30 percent less likely to be under-confident in reading than boys. This pattern is consistent with evidence that adults are more likely to be over-confident in stereotypically gender-congruent domains (Coffman, 2014; Coffman et al., 2019; Bordalo et al., 2019; Shastry et al., 2020).

One key concern with our measures of over- and under-confidence is that they may just capture children’s private information about their own ability, driven by measurement error in the cognitive tests. We have several key pieces of evidence against this concern. First, the math assessment that we use has high test-retest reliability (Hicks and Bolen, 1996). Second, over- and under-confidence strongly persist between waves of the child survey among the 60 percent of our sample with multiple measurements, so our measures seem to capture a stable psychological trait. Next, as we’ve noted, our measures show gender variation that is consistent with prior work on gendered patterns in belief updating, and which we would not expect to see in random testing error. Finally, our results largely persist when we use alternate measures of childhood over- and under-confidence that are less vulnerable to measurement error; we calculate these measures based on test scores and self-reported ability averaged over two waves of the PSID child supplement.

While these four pieces of evidence strongly suggest that our measures of over- and under-confidence capture more than random measurement error on the cognitive test, they do not negate the possibility that children have private information on a form of math ability that the test systematically excludes. This unobserved ability could only explain our results if it is disproportionally weak among girls and correlates with long-run outcomes beyond its association with our many covariate controls (including working memory, general confidence, and family income).

Our main analysis is simple: we estimate the associations between biased beliefs about one’s

---

4Using weights that adjust our sample to be nationally representative, these ranges are 6-30 percent and 6-15 percent, respectively.
math ability in childhood and later educational and employment outcomes, controlling for child-
hood math and reading score decile fixed effects, working memory, general confidence, and a host
of information on respondents’ demographics and family backgrounds.

Children’s biased beliefs in math strongly predict many of their medium- and long-term ed-
ucational and employment outcomes. First, confidence has large associations with educational
achievement: over-confident children score higher than others with comparable prior scores on
math assessments five years later, while under-confident children score lower. Childhood biased
beliefs in math also predict educational attainment: over-confident children are more likely to grad-
uate from high school and under-confident children are less likely to graduate from college than
others with comparable childhood scores. Under-confident children are also less likely to major in
STEM during college, and they attend less selective colleges, though the latter result is imprecisely
estimated. Finally, childhood math confidence predicts key employment outcomes at ages 26 and
up. Under-confident children are less likely to work in STEM occupations as adults, and we find
suggestive evidence that more confident children earn more and are less likely to be unemployed.

While we do not claim that these associations are causal, we do show that they are robust to
several key potential confounders. First, children may form inaccurate beliefs about their ability
in part because of how their parents or teachers perceive them, and these adult beliefs may them-
selves affect children’s later success. One might worry, for example, that our results are driven
by the “Pygmalion effect” in classrooms, where higher teacher expectations for children’s success
improve their later performance (Papageorge et al., 2018; Jussim and Harber, 2005; Wang et al.,
2018). However, we show that our main results are robust to controlling for parent and teacher
expectations for children’s later educational attainment, teacher perceptions of children’s compe-
tence, and parent-reported measures of investment like often doing homework with their child.

Second, children may be reporting their self-perceived ability relative to their school or class-
room, while we evaluate their demonstrated ability relative to a national sample. While we are
limited in our ability to measure school quality, the measures we do have – proxies for school
income, investment, and average achievement – do not correlate with over- and under-confidence,
conditional on our other controls. Controlling for these measures of school quality does not change our results. Finally, our results are also robust to controlling for childhood “Big Five” personality traits, suggesting that over- and under-confidence in math are distinct from these more commonly-studied attributes.\(^5\) In addition to testing these confounders, we also show that our results hold when we use fourteen different formulations of over- and under-confidence – varying all of the key decision points in constructing our main measures – as our key independent variables.

There are two dynamic patterns that could underlie the associations we estimate. First, children’s over- and under-confidence could alter early patterns of educational investments by parents, teachers, or children themselves. Gradually, these differential investment patterns could accumulate into differences in performance and attainment at each juncture in a child’s education, ultimately explaining the gaps that we observe. On the other hand, if children’s biased beliefs persist, they may have direct psychological effects on choices and performance at each stage in a young adult’s development, conditional on his or her performance up to that point. Our evidence suggests that this latter explanation may play a role in the associations we observe. Over- and under-confidence persist through adolescence and into young adulthood (ages 18-27), so biased beliefs could continue to directly affect young adults’ decision-making as they age. We also find that childhood confidence continues to substantially predict later-life education when we hold fixed all intermediate outcomes, though prior education explains much of the relationship between childhood math confidence and labor market outcomes.

Our results suggest that over- and under-confidence merit study as psychological traits with key economic implications. While our results are not causally identified, they are consistent with childhood confidence having important effects on later-life outcomes. Our evidence is also consistent with the idea that those with more confidence fare uniformly better: under-confident children have worse outcomes than their peers with comparable test scores, while over-confident children

\(^5\)The PSID child assessments do not include standard psychometric scales for the Big Five, so we construct proxies for these traits using parents’ reports of children’s behavior and personality. See Appendices B and D for details.
have better.\footnote{Since girls are more likely to be under-confident in math and less likely to be over-confident, these associations could help to explain key gender gaps in the labor market. Unfortunately, our results are too imprecise for us to conclude whether controlling for biased beliefs in math reduces the gender gaps in adolescent test scores, majoring or working in STEM, or earnings.} Our results leave ample room for future work: to experimentally test the impacts of childhood biased beliefs, to clarify the mechanisms underlying the associations we observe, and to design and test interventions that build confidence in childhood and later-life.

Our paper contributes to three literatures in economics.\footnote{Psychology research on academic confidence studies how these beliefs develop as children age (e.g. Eccles et al. 1984) and depend on social constructs like gender and race (e.g. Herbert and Stipek 2005; Usher and Pajares 2006). This work relies on self-reported psychometric scales and does not compare self-reported ability to a measure of objective ability, as we do in this paper.} First, we add to a recent economics literature estimating the returns to psychological or social attributes in the labor market; we provide the first evidence on the returns to over- or under-confidence in the specific academic domain of math. In addition to the work on general self-esteem and long-term outcomes that we cite above, parallel literatures examine the associations between economic outcomes and the Big Five personality traits (Almlund et al., 2011; Heckman et al., 2019), competitiveness (Buser et al., 2021), and children’s time, risk, and social preferences (List et al., 2021). While our data do not measure children’s competitiveness or time and risk preferences, our results are robust to controlling for measures of the Big Five traits in childhood. Together with this prior work, our paper suggests that future work should disentangle the economic importance of these various traits.

Second, we extend the literature on asymmetric belief updating in adults by documenting over- and under-confidence in a large sample of children outside of a lab setting. This heterogeneity mostly matches the lab-based economics literature, which has found mixed patterns of asymmetric updating: some studies find that people over-weight positive signals on average, others find over-weighting of negative signals on average, and some find no average asymmetry (Benjamin, 2019; Zimmermann, 2020). As we’ve noted, the gender gaps in math confidence that we observe are consistent with lab-based evidence that people over-weight positive ability signals in stereotypically gender-congruent domains (e.g. Coffman et al., 2019).

Finally, the studies most relevant to our own examine the role of beliefs about ability in edu-
cational settings. Owen (2020) shows that male college students over-estimate their own ability in STEM and under-estimate the ability of others, while women are more likely to over-estimate others’ ability; informing students about their ability then shrinks gender gaps in beliefs and STEM credits. We find that even children have biased beliefs about their own abilities, with similar gendered patterns. Since children’s beliefs may be more malleable than those of college students, our work suggests that interventions like Owen’s may be fruitful at younger ages. Owen does not assess whether the de-biasing intervention has effects beyond the same semester, but our results suggest that longer-term effects could be substantial.

While Owen (2020) intervenes specifically to change students’ beliefs about their ability, other interventions target self-perceptions more broadly. For example, several studies show that building children’s generalized self-efficacy and grit can narrow gender gaps in both confidence and willingness to compete in math (Falco et al., 2010; Alan and Ertac, 2019). Similarly, Carlana et al. (2018) find that a multifaceted career-counseling intervention among high-achieving immigrant students in Italy increases self-efficacy and successfully closes native-immigrant gaps in pursuing a more academic high-school track. In contrast, we study math-specific confidence.\(^8\) We also show that math confidence predicts long-term outcomes even when controlling for general confidence, so interventions to close math confidence gaps may be important complements to interventions that build general self-efficacy or grit.

Finally, Diamond and Persson (2017), the only related paper to consider long-term outcomes, show that receiving an undeservedly marked-up grade on a test at ages 14-16 leads to higher later \(^8\)Contemporaneous work by Anaya et al. (2021) uses the same data from the PSID and its child supplements to examine the relationship between majoring in STEM and early childhood achievement, self-assessed ability, and parent occupation, though they focus on including parent occupation as a novel explanatory variable in this regression. Like them, we include indicators for whether children’s parents work in STEM in our main specifications, but adding these controls does not change our results. Anaya et al. also describe similar patterns of gender differences in beliefs about ability as we do, but they do not specifically study over- and under-confidence or their relationships with long-term outcomes. In addition to this difference in our central research questions, we see our work as building on theirs in three ways: (1) we use a more comprehensive set of available data from the PSID and its child supplements; (2) we consider a much larger set of outcomes observed over a much longer time frame; and (3) we define several new measures of over- and under-confidence to deal with complications with the raw data, an issue that Anaya et al. do not discuss.
test scores, more likely high school and college graduation, and higher earnings. Since marked-up scores in one subject raise later scores across all subjects, the authors argue that these effects arise in part by changing students’ beliefs about their own ability. However, they do not actually observe students’ beliefs about their own ability, as we do. Together, our papers strongly suggest that students’ biased beliefs about ability matter for later educational and employment outcomes.

Finally, our work differs from all of this prior research in that we separately study over- and under-confidence, though our results ultimately suggest that more confidence improves outcomes across the board.

The paper proceeds as follows. Section 2 lays out a conceptual framework for how childhood confidence might affect economic outcomes, and Section 3 describes our sample and our measures of biased beliefs. Section 4 analyzes the prevalence and predictors of childhood over- and under-confidence in our sample, and Section 5 describes our strategy for estimating the links between biased beliefs in math and long-run economic outcomes. Section 6 presents results, Section 7 describes the stability of our results to potential confounders and alternate definitions of confidence, and Section 8 explores the dynamic patterns that these long-term associations might follow.

2 How might childhood math confidence affect economic outcomes?

Ability or skill is a primary independent variable in almost every economic model of student and worker decision-making. These include settings where agents are investing in their own futures, like deciding to continue with schooling, choosing a college major or career, or searching for a job (e.g. Becker, 1964; Roy, 1951; McCall, 1970; Borjas, 1987; Kirkeboen et al., 2016), as well as settings where teachers or parents decide, for example, how to invest in or tailor their pedagogy to a child (Fryer, 2018; Dizon-Ross, 2019).

Over- and under-confidence enter any of these models if ability is imperfectly observed: by parents, teachers, and even by the student or worker themself. Replace ability in these standard models with perceived ability, and the potential effects of over- and under-confidence become clear. In cases where ability and effort are complements, like college applications, over-confident agents may work harder. In cases where ability and effort are substitutes, like some school tests, over-
confident agents may reduce their effort. Bénabou and Tirole (2002) model how over-confidence can persist in equilibrium in either setting. Consistent with the case of complementary effort and ability, psychological theories of motivation, including Bandura’s (1986) Social Cognitive Theory or Expectancy-Value Theory (see Wigfield and Eccles (2000)), emphasize that individuals are more likely to attempt tasks in which they feel competent, work harder on those tasks, and are more likely to succeed.

Over- and under-confidence will also affect outcomes in any setting where teachers or parents decide how to invest time and resources into children based on their perceptions of each child’s ability. If adults interpret more confident children as more skilled, they may over-invest in over-confident children and under-invest in under-confident children. Dizon-Ross (2019) shows that parents have inaccurate beliefs about their children’s academic performance, and that correcting those beliefs causes them to adjust their investments. Similarly, Papageorge et al. (2018) show that having a teacher with higher expectations increases a student’s chance of completing college. The same forces could operate in job applications, where potential employers are uncertain about applicants’ skill: Schwardmann and van der Weele (2019) show that interviewers rate more confident job applicants more favorably.

**Our focus on confidence gaps in math, not reading**

Our data includes all of the necessary information to identify over- and under-confidence in both math and reading, but the remainder of the paper will focus just on biased beliefs in math. We make this choice for several reasons. First, performance in math can be measured more objectively than performance in reading, so children’s beliefs about their math ability may be more precise. Next, multiple strains of research suggest that math ability during childhood and young adulthood more strongly predicts later achievement than does reading ability (e.g. Duncan et al., 2007; Castex and Kogan Dechter, 2014; Goodman, 2019). We find similar patterns in our data in Appendix Table A1, where we regress our main education and employment outcomes on childhood test scores and the set of controls that we will use throughout our main analysis. While both CDS
math and reading score percentiles predict later academic achievement and attainment, only math scores predict earnings, unemployment, and majoring in STEM. Thus, we would expect children’s perceptions of their own ability in math to also more strongly link with later-life achievements than their self-perceptions in reading. Finally, the Bureau of Labor Statistics predicts that employment in STEM occupations will continue to grow at faster rates than non-STEM occupations through 2030, so math ability may become an even more important predictor of success in the labor market.

That said, we do conduct all of the subsequent analysis for reading confidence in addition to math. As expected, reading confidence robustly predicts few educational or employment outcomes. We present results that mirror our main analysis for reading in Appendix Tables A2-A5.

3 Measuring confidence and later-life outcomes in the PSID

3.1 Sample and survey design

We explore the links between biased childhood beliefs and outcomes in young adulthood using the rich data system of the Panel Study of Income Dynamics (PSID). The PSID was first collected in 1968 among a sample of 5,000 nationally-representative households. It has since surveyed the descendant households of the original sample annually from 1968 to 1997 and biennially thereafter, adjusting the sample in 1997 to again make it nationally representative.

We combine the core PSID with two supplements that follow respondents from childhood into young adulthood: the Childhood Development Supplement (CDS) and the Transition into Adulthood Supplement (TAS). The CDS was introduced in 1997, sampling up to two children per PSID household who were then between the ages of 0 and 12 (3,563 children). The CDS collects detailed information from children themselves, from their primary caregivers, and from their elementary school teachers on areas including children’s cognitive and emotional development, health, and exposure to parenting practices. The original CDS sample was re-interviewed in 2002-2003, then aged 5-17, and those still below age 18 were included in a third CDS wave in 2007.

In 2005, the PSID introduced the TAS as a bridge between the CDS and the main PSID survey for CDS respondents, the oldest of whom had reached ages 18 to 20 by that year. The TAS has been collected biennially since 2005, with younger CDS respondents aging into the TAS sample.
at 18. Individuals participate in the TAS until they become economically-independent heads of their own household, at which point they enter the adult PSID sample and are surveyed every two years. The TAS is designed to capture respondents’ social and career development as they enter adulthood; we use its modules on education, employment, income, and personality.

The PSID-CDS-TAS data structure is uniquely suited to exploring the links between childhood confidence and long-term educational and employment outcomes. First, the CDS both administers a math test and asks children to evaluate their own math ability; we combine children’s test scores and self-assessments to identify over- or under-confidence in math. Section 3.2 below will detail the CDS tests, self-assessments, and our key confidence measures. Second, following CDS children into the TAS and then the PSID allows us to observe detailed data on educational and employment outcomes over 22 years, following our oldest respondents into their mid thirties. Using restricted data from the TAS, we are also able to link respondents with college quality data from the first college they attended in the first year they attended that college. We construct college quality measures using data from the National Center for Education Statistics (NCES). Finally, the extensive data on parents’ employment and income in the PSID and on parenting practices and other child characteristics in the CDS allows us to control for many covariates that could confound the relationship between biased beliefs and long-run outcomes.

In particular, the detailed child module in the CDS allows us to control for other forms of ability and confidence that are distinct from skill and confidence in math, but which may correlate with them. We construct a measure of general confidence as the mean of standardized variables capturing whether children see themselves as broadly competent (see Appendix B for details); we have no measure of true ability by which to normalize this general confidence scale, so it is an ambiguous mixture of over- or under-confidence and unobserved ability. Thus, we use general confidence as a control for unobserved abilities and other dimensions of confidence that may correlate with biased beliefs in math and also affect later-life outcomes. We also control for children’s scores on the Digit Span subtest of the Wechsler Intelligence Scale for Children (Revised), a measure of short-term memory.
The CDS and core PSID also collect detailed household information on total family income, household heads’ education, primary caretakers’ values and mental health, household structure, and financial characteristics like whether the household receives food stamps. Finally, we use restricted data from the CDS to link students with elementary and middle school quality data from the NCES and the Stanford Educational Data Archive (SEDA). Section 5 will outline in detail the family and child variables that we control for in our main analysis.

Our final sample consists of the 2,985 CDS respondents with at least one year of math cognitive tests and self-assessments in the CDS, about 84 percent of all CDS respondents. Summary statistics for this sample are reported in Appendix Table A6; all variables are taken from the CDS or PSID survey from the same year in which we first observe childhood over- or under-confidence in math. We generate many of these variables by standardizing multiple related questions and taking the average to form indices (see Appendix B for variables that make up each index).

Our sample is non-randomly selected from the national population, both because the initial 1968 PSID sample oversampled low-income families and because there is unobserved selection in whether CDS participants report math test scores and self-assessments. This selection appears in our sample statistics in notable ways. First, our sample is disproportionately Black: 45.8 percent of our sample are White, 41.7 percent are Black, and only 7.5 percent are Hispanic, while the U.S. Census Bureau reports that 69.1 percent of the US residents were White, 12.1 percent Black, and 12.5 percent Hispanic in 2000 (Greico and Cassidy, 2001). While the Census Bureau reports median household income in 1997 of $55,336, our sample’s median income is slightly lower, at $52,029 (both in 2016 USD). On the other hand, our study sample performs disproportionately well on the CDS standardized tests: we observe median CDS math and reading score percentiles

---

9 Most children who are missing test scores or self-assessments lack this data because they skipped the entire section of the CDS administered to the child, while completing the survey portions administered to the primary caregiver. These respondents largely have similar demographics to those for whom we observe confidence measures, but their mothers are less likely to have a high school degree, they have lower total family income, and they are about a year younger. Students who take the math cognitive assessments but do not give self-assessments (about 25 percent of the children who are missing test scores or self-assessments) score much lower on both the math assessment and the Digit Span memory test (Appendix Table A18).
of 60 and 54, respectively, relative to national norming samples.

While we do not weight our sample to be nationally representative in our main analysis, we include results that do so in Appendix Tables A7-A10. These weights are based on those published by the CDS, which capture the inverse probability of respondents’ inclusion in the CDS sample; we then recalibrate these CDS weights via iterative proportional fitting, or raking, to ensure that our sample matches marginal distributions of percentile CDS math scores, race in 2000, and total household income in 1997. As expected, our results are less precisely estimated when we use weights, though they remain qualitatively similar. These recalibrated weights particularly underweight children with high CDS math scores, in some cases leaving us underpowered to detect the correlations between under-confidence and long-term outcomes.

3.2 Measuring over- and under-confidence in math

Data on children’s self-reported and demonstrated ability in math

The CDS assesses children’s math skills using the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R), a common test of academic achievement used by school psychologists in the 1990s (Stinnett et al., 1994; Hicks and Bolen, 1996; Duffy and Sastry, 2014). The CDS administers the Applied Problems subtest of the WJ-R, comprising 60 word problems of increasing difficulty that assess math reasoning and knowledge. Each child completes only a subset of the test, beginning at a “basal” level, where they answer six consecutive questions correctly, and ending at a “ceiling” level, where they get six consecutive questions wrong. The CDS then reports each respondent’s percentile rank relative to the nationally-representative WJ-R norming sample for their age group; we use these percentile ranks as our measure of each child’s demonstrated ability in math. Panel A of Figure 1 shows the distribution of these scores in our sample.

In addition to collecting this measure of performance in math, the CDS also asks all respondents ages 8 or older to assess their own ability in math, asking them to answer “How good at math are you?” on a scale of 1 (not at all good) to 7 (very good). Children never receive their scores on the

---

10 The CDS test also included 58 WJ-R questions on calculation skills in the 1997 wave, and we use this test in the next section to assess the reliability of our over- and under-confidence measures.
WJ-R math test, so these self-reports do not reflect feedback from the CDS. Panel B of Figure 1 shows the distribution of these self-assessments. Math self-perceptions are highly skewed towards positive responses, with over 89 percent of respondents ranking themselves as “Okay” or better at math. This skew may be partially explained by the distribution of percentile scores in Panel A, which skews heavily towards higher-performing children. While shifted upwards, children’s self-reports do contain information about objective ability: in Panel C of Figure 1, average math test percentiles rise almost linearly with self-reported ability in math.

We measure children’s over- and under-confidence in math in the first wave of the CDS in which they have non-missing cognitive test scores and self-assessments, leaving us with a sample of 2,985 children.\textsuperscript{11} We first measure confidence for the median child before age 11, and we observe confidence by age 13 for almost 90 percent of children. Thus, we will interpret our measures as childhood over- and under-confidence in math.\textsuperscript{12} Throughout, our analysis will control for both birth year and the age at which we first observe confidence.

**Defining binary measures of over- and under-confidence**

We first identify over- and under-confidence in math using large mismatches between children’s score percentiles and their self-assessments. In particular, we classify any respondent as under-confident in math if she scored above the 75th percentile nationally and ranked her own ability at 1 to 4, corresponding to the bottom 47 percent of the subjective ability distribution in our sample, or if she scored above the 50th percentile nationally and ranked herself at 1 to 3, corresponding to the bottom 10 percent of the subjective ability distribution. We define over-confidence among low-achievers using similar thresholds, but we account for the skewed self-assessment distribution by using stricter cut-offs to identify biased beliefs. In particular, we identify any respondent as over-confident in math if she scored below the 25th percentile nationally and rated her own ability

\textsuperscript{11}We first observe confidence from the 1997 CDS wave for 1,075 children, from the 2002 CDS wave for 1,347 children, and from the 2007 CDS wave for 563 children.

\textsuperscript{12}In addition to measuring children’s demonstrated and self-assessed ability in math, the CDS also assessed children’s reading performance using the WJ-R subtest for Passage Comprehension and asked children “how good at reading are you?” on a scale of 1 to 7. We use these WJ-R reading scores and self-assessments to define measures of reading over- and under-confidence that exactly parallel our measures of biased beliefs in math.
at 6 or 7, corresponding to the top 39 percent of the subjective ability distribution in our sample, or if she scored below the 50th percentile and rated herself at 7, corresponding to the top 22 percent of the subjective ability distribution.

These measures of math over- and under-confidence have several key strengths: they are easy to define and observe, they refrain from putting too much stock in the cardinal value of children’s self-assessed ability, and they account for the upward skew in self-assessments, which we consider to be a form of response bias separate from over- or under-confidence.

However, our measures also have several limitations. First, we can only identify over-confidence among children scoring below the 50th percentile and under-confidence among those scoring above the 50th percentile; thus, our results examine optimism bias among childhood low-scorers and imposter syndrome among childhood high-scorers. This strategy matches the existing literature, which typically documents imposter syndrome among high-achievers (Sakulku, 2011). Another limitation is that these measures are not directly comparable to measures of over- and under-confidence from the lab-based literature, which can precisely measure respondents’ beliefs about their quiz performance or rank relative to a group (e.g. Coffman et al., 2019; Möbius et al., 2014; Eil and Rao, 2011). Our measures of over- and under-confidence, in contrast, identify coarse categories of children with large gaps between their self-assessments and observed scores. Our second measure of biased beliefs, described below, aims to partially address these limitations.

**Defining a more continuous measure of biased beliefs**

Our second measure of biased beliefs identifies confidence gaps as the difference between children’s self-reports and their observed performance on the CDS math test. To transform these objects to the same scale, we split the distribution of children’s percentile scores uniformly into seven bins, where 1 includes the lowest 14 score percentiles and 7 includes the highest 14 score percentiles relative to the national norming sample. We then assume that students with full information about the national distribution of scores and their place in it would have self-reported their math ability as the bin from 1 to 7 in which their score percentile falls, and we take the difference between their actual self-report and this bin as our measure of biased beliefs. This measure then
takes on values of the integers from -6 to 6. For ease of interpretation, we standardize this variable to have mean 0 and standard deviation 1 throughout the rest of the paper.

This measure has three strengths relative to our main measure: it allows for more granularity in the extent of biased beliefs, aligns more closely with measurements of biased beliefs in the lab-based literature, and relies on fewer choices by the authors. However, by assuming that we can identify even small biases in beliefs about math ability, it is more likely to conflate actual biased beliefs with children’s private information about their math ability (described in more detail in the next section). It may also be confounded by a form of reporting bias – separate from over-confidence – that generates the overall upward skew in self-reports relative to test scores (Figure 1). We present results for all outcomes using both the binary and more continuous formulations of biased beliefs, and in general the results are extremely consistent.

To ensure that our main results do not arise just from our particular choice of confidence measures, we show that our results are robust to using alternate definitions of both our indicators for over- and under-confidence and this more continuous measure of biased beliefs. We describe these alternate measures in Section 7.

3.3 Biased beliefs or measurement error?

One key concern with our measures of biased beliefs in math is that they may conflate over- and under-confidence with children’s private information about their math ability, perhaps driven by measurement error in the WJ-R assessment. Four key pieces of evidence support the claim that our measures truly capture biased beliefs in childhood.

First, prior work has shown that the WJ-R assessment is a reliable measure of children’s math skills, with test-retest reliability for the applied math problems of about 0.85 in large samples (Hicks and Bolen, 1996). We can also verify WJ-R reliability across math domains in our sample using the 1997 wave of the CDS, which was unique in administering both the Calculation subtest of the WJ-R and the Applied Problems subtest. For the 1,450 children who took both tests, the

13Several studies find somewhat lower test-retest reliability for certain ages, at around 0.75, though these studies use small samples (Shull-Senn et al., 1995)
The correlation in percentile ranks on the two sections is 0.69. Our binary designations of children as over-confident, under-confident, or neither is also highly consistent whether we measure objective math ability using children’s percentile scores on the Calculation or Applied Problem subtest: 81 percent of children with both measures are classified in the same category regardless of which ability measure we use. Another ten (nine) percent switch from under-confident (over-confident) to neither or vice versa. When we use our more continuous measure of degrees of confidence, which takes on integer values from -6 to 6, 32 percent of children are assigned the same value regardless of which math test we use as the measure of demonstrated ability, 62 percent are within one integer, and 83 percent are within two integers. See Appendix Figure A1 for the full joint distribution of the more continuous confidence measures based on the two math subtests.

Second, our measures of childhood math confidence persist strongly over time. About 60 percent of the children in our sample appear in two waves of the CDS, allowing us to construct two measures of over- and under-confidence taken five years apart. Children appear in a second CDS wave at ages 13 to 19, so these second-wave measures capture biased beliefs in adolescence. Table 1 regresses our adolescent measures of biased beliefs on our childhood measure of the same variable, controlling for a set of demographics and parent characteristics that we will use throughout our later empirical analysis; we outline these specifications in detail in Section 5. These regressions show substantial persistence: respondents who were over-confident in math as children are about 3 times as likely (12 pp more likely) to be over-confident in math as adolescents, while under-confident children are about 1.7 times as likely (4 pp more likely) to be under-confident as adolescents. Similarly, a one-standard-deviation increase in the degree of biased beliefs in childhood predicts 0.18sd more biased beliefs in adolescence. If our confidence measures just captured random testing variability, we would not expect to see such substantial persistence.

14In addition to the controls that we use throughout our empirical analysis, the regressions in Table 1 also control for our second observation of children’s test score deciles in math and reading, measured at the same time as adolescent confidence. We add these controls to purge any correlations induced by the effects of childhood confidence on adolescent test scores, since childhood over- and under-confidence predict later test scores (see Section 6) and higher-scoring (lower-scoring) children are mechanically more likely to be classed as over-confident (under-confident).
Third, our main results are largely robust to using measures of over- and under-confidence that reduce potential measurement error by combining observations of children’s test scores and self-reported ability across two waves of the CDS; in contrast, our primary results only use children’s first-observed scores and self-reports to identify biased beliefs. We discuss these measures and results in more detail in Section 7. If measurement error is uncorrelated across tests taken 5 years apart, these average confidence measures will be less vulnerable to it than are our main measures.\footnote{While these average measures are less likely to reflect measurement error in the CDS tests, we do not use them as our preferred measure of confidence for three reasons: (i) over- and under-confidence at older ages may be more likely to be confounded by unobserved variables; (ii) we are interested in adolescent test scores and confidence measures as outcome variables; and (iii) only 60 percent of our sample has confidence measurements over multiple waves of the CDS.}

Finally, we describe in the next section that we observe substantial gender gaps in math over- and under-confidence, with girls more likely to be under-confident and less likely to be over-confident. This pattern is consistent with gender stereotypes about math ability, which may shape children’s beliefs even at young ages, and mirrors results for adults in the lab (e.g. Coffman et al., 2019). Our measures of over- and under-confidence could only be entirely explained by measurement error if this error took a similar gendered pattern, beyond its correlation with demonstrated ability on the CDS tests and with the many other controls we outline in Section 5.\footnote{Differential random error by gender could not fully explain the gendered patterns of over- and under-confidence we observe, since the gender with more variable performance would be more likely to be both over- and under-confident. Nonetheless, comparing boys’ and girls’ performance on the Calculations and Applied Problems subtest in the 1997 CDS sample suggests that neither gender has differentially variable test performance. 81\% of both boys and girls receive the same binary confidence designation when calculated using either the Calculations or Applied Problems percentile score as a measure of math skill, and the joint distributions of the more continuous measures are very similar for boys and girls (Appendix Figure A1).}

We consider a few possible sources of non-random measurement error that could generate these patterns: skill in some dimension of math that the test does not cover, test-taking anxiety, and test-taking motivation.

As we’ve noted, an unobserved dimension of math ability could only explain our results if boys perform better than girls on that dimension conditional on their WJ-R applied problems scores. The CDS data allows us to test for gender gaps in one central dimension of math skill that our main test scores do not directly capture: calculation skills. Using the 1997 CDS sample, when children took
both the WJ-R Calculation subtest and the WJ-R Applied Problems subtest, we find no evidence that boys have better calculation skills conditional on the applied problems scores that we use in our main analysis.\(^\text{17}\)

Next, differential measurement error in the CDS math tests could arise if boys or girls are more prone to testing anxiety that impairs performance. While past work finds that boys show higher cortisol reactivity, a measure of physiological arousal, during test-taking (Weekes et al., 2006; Stroud et al., 2002), other research suggests that physiological arousal only impairs performance when students psychologically appraise it as an indicator of potential failure (Jamieson et al., 2013; Mattarella-Micke et al., 2011). Girls tend to have higher psychological test anxiety (worrying about failing, forgetting everything, etc.) and math anxiety (tension, dread, or fear around math), and most commentary suggests that it is these psychological manifestations of anxiety that pose first-order risks to test performance (Devine et al., 2012; Erturan and Jansen, 2015; Ballen et al., 2017). Thus, we would expect girls’ test performance to differentially lag their true skill, producing gender gaps in confidence that would conflict with our empirical results.

Finally, we turn to test-taking motivation. Past work finds that girls are somewhat more motivated than boys to exert effort on low-stakes tests, so boys’ CDS math scores may be differentially low relative to their true skill in math (Segal, 2012; DeMars et al., 2013; Gneezy et al., 2019). Then, boys may appear more over-confident by our measures. While it is hard to fully eliminate this possible confounder in our setting, our results are robust to controlling for agreeableness and conscientiousness, two Big-5 personality traits that are positively correlated with unincentivized test effort (DeMars et al., 2013; Segal, 2012). (See Section \textit{7} for more details.)

Together, most evidence from our empirical setting and from past work on test-taking strongly suggests that our confidence measures capture a meaningful psychological trait. However, we cannot fully eliminate the risk that these measures capture children’s private information on some

\(^{17}\)We estimate the following regression: $\text{CALC} \text{ptile}_i = \beta_0 + \beta_1 \text{AP} \text{ptile}_i + \beta_2 \text{Female}_i + \beta_3 \text{AP} \text{ptile}_i \times \text{Female}_i + \epsilon_i$. Coefficient $\beta_3$ is not significantly distinguishable from zero, and $\beta_2$ is significant and positive. Thus, girls have stronger calculation skills than boys conditional on their Applied Problems scores, which would tend to make girls look more over-confident by our measures, the opposite of what we find.
aspect of math ability that the test systematically excludes. Any such confounder could only explain our results if it is differentially weak among girls and affects outcomes beyond its correlation with demonstrated math ability, general confidence, digit span score, reading ability, and the many other controls we outline in Section 5.

4 Patterns of over- and under-confidence in the population

This section documents the prevalence and correlates of over- and under-confidence in our sample. Besides documenting biased beliefs in math in a real-world setting, these results are useful both to validate our measures of biased beliefs and to inform our strategy for estimating the links between childhood confidence and long-run outcomes, which we describe in Section 5.

4.1 Prevalence of biased beliefs

First, we find substantial over- and under-confidence among children in our sample: using our main binary measures, 8.5 percent of children are over-confident at their first measurement, while 12 percent are under-confident.\footnote{We find similar results when applying our raked weights to obtain nationally representative estimates: 9.2 percent of children are over-confident and 9.7 percent are under-confident.} Since this measure identifies over- and under-confidence as large gaps between children’s self-assessed and objective performance, these shares are strikingly high. Turning to our more continuous measure of biased beliefs, 20 percent of children report the same bin as their percentile score would imply, 8.7 percent of children report ability levels that are at least 3 bins lower than that of their score, and 17 percent report ability levels that are at least 3 bins higher, where each bin spans 14 percentiles of observed scores. See Appendix Figure A2 for the full distribution of the continuous confidence measure. It is notable that over- and under-confidence are both prevalent in this large sample, given psychology’s focus on over-confidence (Moore and Healy, 2008) and the mixed evidence from lab experiments on asymmetric belief updating (Benjamin, 2019).

Second, older children have more accurate beliefs. Panel A of Appendix Figure A3 plots the share of children who are over- or under-confident in math by age; Panel B plots the cumulative
density function for the continuous confidence measure for three age groups, pooling respondents’ observations across CDS waves. Both panels show that children are most likely to have incorrect beliefs about their math ability when they are young, and average belief accuracy increases almost monotonically as children age. We focus on the associations between confidence and later-life outcomes using first-observed confidence (observed by age 13 for 90 percent of the sample), so our confidence observations are drawn from young ages with more biased beliefs. We will eliminate bias due to the timing of our confidence measurements by including fixed effects for the age at which confidence was measured in all regressions.

4.2 Biased beliefs and other child characteristics

Over- and under-confidence correlate with other child characteristics in largely expected ways (Table 2 and Appendix Table A11). First, and unsurprisingly, children with higher general confidence are more likely to be over-confident and less likely to be under-confident in math, and children with higher digit span scores are less likely to be under-confident. Math test score deciles strongly predict confidence gaps (though some of this correlation arises mechanically from how our measures are constructed), while reading test score deciles do not (in Appendix Table A11). To avoid conflating biased beliefs in math with other skills or general confidence, we will control for children’s general confidence, digit span scores, and test score decile fixed effects in math and reading in all regressions of later-life outcomes on childhood biased beliefs.

Conditional on these measures of ability, children who have ever been in a gifted program are 8.7pp less likely to be under-confident in math and 2.6pp more likely to be over-confident. These correlations could reflect that schools and children share private information on children’s ability conditional on CDS scores, that being in a gifted program alters children’s confidence, or that children’s confidence influences their treatment at school conditional on ability. To avoid controlling for mediators of the effects of confidence, our regressions will not control for this variable or other signals of ability from schools, like repeating a grade.19

19Math over- and under-confidence also correlate with children’s other attitudes towards math and school in reasonable ways (Appendix Table A19), suggesting that our measures isolate over- and under-confidence in the particular domain of math. See Appendix C for more discussion.
Finally, gender is the strongest demographic predictor of math confidence. Girls are 2.3pp (17 percent) more likely to be under-confident and 2.7pp (27 percent) less likely to be over-confident in math than boys with the same score deciles, and on average, girls’ biased beliefs are 0.1 standard deviations (sd) lower than the average boy’s. Note that girls do not have more accurate beliefs, simply more negatively-biased ones. This finding is consistent with prior literature showing that adults are more over-confident in gender-congruent domains (e.g. Coffman et al., 2019), but it is notable that we find it in children, the majority of whom have not yet entered puberty. These gender differences are present at almost all ages, but due to small sample sizes the patterns are imprecise (available upon request).

Perhaps surprisingly, we find no significant links between children’s math confidence and their parents’ education or occupation, household income, or race conditional on all other characteristics. We note, though, that noise in these estimates means we cannot reject potentially large correlations.

5 Confidence and long-term outcomes: Empirical strategy

Our empirical strategy is simple: we estimate the associations between biased beliefs in math ability and later education and work outcomes, holding fixed measured childhood ability. We use the PSID’s rich data on childhood environment and family characteristics to control for extensive

\footnote{Appendix Figure A17 plots the raw data underlying our confidence measures (analogous to Figure 1) separately for boys and girls.}

\footnote{In fact, there is no gender gap in the likelihood of having accurate beliefs (degrees of over- and under-confidence equal to zero) or almost accurate beliefs (degrees of over- and under-confidence between -1 and 1). Results are available upon request.}

\footnote{Appendix Figure A18 shows that this gender gap is extremely robust to using alternate definitions of over- and under-confidence and alternate ways of calculating the more continuous degrees of confidence measure. The figure plots the coefficient on the female indicator when we exchange the dependent variables in Table 2 with these alternate measures (discussed further in Section 7).

\footnote{Given the substantial gender gaps in the prevalence of over and under-confidence in math, we also look further into the role of gender by testing whether childhood gender gaps in math under-confidence explain the gender gaps that exist in some later education and employment outcomes: adolescent test scores, majoring in STEM, and earnings. Specifically, we estimate the change in the coefficient on gender when we estimate our preferred specification with and without the indicator for under-confidence (following Buser et al. (2021)). The results (available upon request) are extremely noisy, so we leave it to future research to determine whether over- and under-confidence in math can help explain these and other gender gaps.}
pre-determined confounders, but we refrain from interpreting our estimates as the causal effects of confidence. We estimate the following specification:

\[ Y_{it} = \alpha + \beta_1 Over_{i0} + \beta_2 Under_{i0} + A_{i0}' \mu + X_{i0}' \delta_1 + X_{i0}' \delta_2 + \gamma_s + \omega_t + \varepsilon_{it} \]

where \( Y_{it} \) is individual i’s outcome of interest in adolescence or adulthood, measured in wave t of the CDS, TAS, or PSID, and \( Over_{i0} \) and \( Under_{i0} \) are indicators for being over- or under-confident in math as a child, respectively. All of our main tables also include regressions in which we replace \( Over_{i0} \) and \( Under_{i0} \) with the single \( ZConf_{i0} \) variable, which captures the degree to which a child is over- or under-confident in standard deviations. Due to power limitations, we assume that \( ZConf_{i0} \) has a linear relationship with our outcomes of interest.\(^{24}\)

Next, all of our regressions include \( A_{i0} \), a vector of controls for childhood ability. In particular, \( A_{i0} \) includes linear controls for childhood digit span score and general confidence, as well as fixed effects for test score deciles in both reading and math.\(^{25}\) Our basic specification also includes state fixed effects \( \gamma_s \), TAS or PSID wave fixed effects \( \omega_t \),\(^{26}\) a set of child controls \( X_{i0}^C \), and a set of parent controls \( X_{i0}^P \). In our first specification, \( X_{i0}^C \) and \( X_{i0}^P \) include only variables that are certainly

\(^{24}\)Appendix Figures A19, A20, and A21 show our main results when we relax this assumption; we plot the coefficients on indicators for each integer value of the variable underlying \( ZConf_{i0} \):

\[ Conf_{i0} = -6, \quad Conf_{i0} = -5, \quad ..., \quad Conf_{i0} = 6. \]

While these results are noisy, taking the point estimates at face value suggests that this linearity assumption is reasonable. We also show in Section 6.5 that we cannot generally reject the null hypothesis that over- and under-confidence predict economic outcomes in similar (opposite-signed) ways, further supporting this linearity assumption.

\(^{25}\)One might worry that controlling for general confidence absorbs too much of the variation in math over- and under-confidence if over- or under-confidence in math is a dimension of confidence in general. While the economic impacts of general confidence are certainly of interest, we take the conservative approach of isolating math-specific over- and under-confidence as cleanly as possible by including controls for general confidence. That said, our results are remarkably similar with or without the control for general confidence (available upon request).

\(^{26}\)For some outcome variables, like earnings and unemployment, we have multiple years of outcomes across TAS and PSID waves for each respondent. In contrast, we observe our educational outcomes (e.g. whether respondents ever majored in STEM) only once per respondent; we do not include survey wave fixed effects in regressions linking childhood confidence to these outcomes. Note that we do not include respondent fixed effects even in regressions with multiple outcome observations per respondent, since we only measure childhood confidence once. Note that we cluster standard errors by family in all regressions.
unaffected by respondents’ childhood math confidence. Specifically, $X_{i0}^C$ includes fixed effects for race, birth year, quarter of birth, gender, and age at which we observe confidence, and $X_{i0}^P$ includes family income, its square, and fixed effects for both parents’ levels of education. All variables indexed at $t = 0$ are from the first CDS wave in which a child had WJ-R scores and an ability self-assessment. Since about two-thirds of the children in our sample have a sibling in the sample, we cluster standard errors by family. Our coefficients of interest are $\beta_1$ and $\beta_2$.

Our second specification takes advantage of the detailed interviews of children’s caregivers in the CDS to add additional controls for child and family characteristics that may correlate with both confidence and long-run outcomes. In addition to expanding the set of child controls, $X_{i0}^C$, with the primary caregiver’s assessment of the child’s general health, this specification supplements $X_{i0}^P$ with controls for whether the family receives government transfers, whether the father is in the household, whether the household has two adults, and parents’ beliefs about gender norms and the qualities that are most important for success. Specification two also adds controls for parent confidence and mental health, including the primary caregiver’s Rosenberg self-esteem score, Pearlin self-efficacy score, and “aggravation-in-parenting” score compiled by the PSID (see Appendix B for details). Finally, we add four indicators for whether the child’s mother and father work in STEM or another high-education occupation (based on the findings of Anaya et al. (2021); see footnote 8). We will treat this model as our preferred specification throughout the text, but results are generally consistent across these two specifications.

6 Confidence and long-term outcomes: Results

The following section documents strong associations between childhood under- and over-confidence in math and key later-life outcomes: adolescent test scores, whether respondents graduated from high school and college, college major, career choice, earnings, and unemployment. We present these main results in Tables 3, 4 and 5.
6.1 Medium-term educational achievement

We first examine the links between childhood confidence and medium-term educational achievement, which we measure using adolescent scores on the CDS math assessments. We observe these scores at children’s second observation in the CDS, about 5 years after we first observe their confidence in math.

Children’s biased beliefs in math significantly predict adolescent math performance (Table 3, columns 1 and 2). Using our binary measures (Panel A), children who are over-confident in math score 2.6 percentiles (standard error = 1.5p) higher on the math assessment five years later than others with comparable baseline scores, while under-confident children score 5.8 percentiles (se = 1.5p) lower. Using our more continuous measure (Panel B), we find that a child with 1 standard deviation (sd) higher math confidence in childhood scores 2.8 percentiles (se = 0.57p) higher on the math assessment 5 years later than others with comparable baseline scores. Over-(Under-) confident children differ from children without biased beliefs in these binary metrics by an average gap of 1.9sd (-1.5sd) in more continuous degrees of confidence, so our estimate magnitudes are remarkably consistent across the two panels. In contrast, there is no relationship between childhood math over- or under-confidence and adolescent reading scores using either measure of biased beliefs (Table 3, columns 3-4).

These associations are large relative to the links between raw math ability and later scores. Increasing one’s childhood math score by 10 percentiles is associated with scoring on average 5.5 percentiles higher in adolescence (coefficients on decile fixed effects are estimated in the same regression and can be found in column 1 of Appendix Table A12). Thus, being over- (under-) confident in math predicts as large a gap in adolescent test scores as does increasing (decreasing) one’s childhood math test score by 5-10 percentiles.

6.2 Educational attainment

Besides predicting later test scores, biased beliefs in math during childhood also predict important educational outcomes down the line. We first consider high school and college graduation.

Children who are over-confident in math are 6.2 percentage points (se = 2.6pp) more likely to
graduate from high school, and children who are under-confident in math are 5.4pp (se = 2.8pp) less likely to graduate from college (Table 3, columns 5-8, Panel A). Since only 30 percent of our sample graduates from college, being under-confident predicts an 18 percent drop in the likelihood of attending college. We find very similar results using our more continuous measure in Panel B: a child with 1sd higher math confidence in childhood is 1.8pp (se = 1.0pp) more likely to graduate from high school and 3.2pp (se = 1.1pp) more likely to graduate from college, though the first is only marginally significant. Again, the magnitudes of these results are similar regardless of which confidence measure we use. These gaps are large relative to the associations between childhood math scores and educational attainment in our data: childhood math scores generally do not predict high school graduation, and increasing test scores by one decile is associated with being 2.8pp more likely to graduate from college on average (Appendix Table A12, columns 3-4).

6.3 College quality, college major choice, and post-college education

Next, we consider later-education outcomes among those who went to college: college quality, college major choice, and whether respondents complete a graduate degree. Since we restrict to college graduates, these regressions use much smaller samples than for our previous outcomes.

First, we find imprecise links between childhood math confidence and the quality of colleges that children later attend. We consider two quality measures: first, an index of general college quality, and second, colleges’ 75th-percentile math SAT scores among incoming freshmen – a more specific measure of math quality.27 We focus our discussion on colleges’ 75-percentile math scores (Table 4, column 3 and 4), but our results are similar using the more general college quality index (columns 1 and 2). Under-confident children attend schools whose 75th-percentile math SAT scores are 10.8 points (se = 5.9 points) lower than others with the same childhood scores (p = 0.07); with 95 percent confidence, we can reject that under-confident children attend schools with

27 Following Cohodes and Goodman (2014), we construct an index of college quality by taking the first component from a principal component analysis of colleges’ 75th-percentile math SAT scores among incoming freshmen, graduation rates, and per-pupil instructional expenditures, separately by year. Details on variable construction are available in Appendix B. We then standardize this index to have mean 0 and standard deviation 1 in the full sample of four-year colleges in the US by year.
math SAT scores that are over 0.7 points higher or 22.4 points lower. Over-confidence is not significantly associated with college quality among childhood low-scorers, but again we observe wide confidence intervals: in our preferred specification, we cannot reject that over confident children attend colleges that have 20 points lower to 26 points higher SAT scores than their peers. Using our more continuous measure of biased beliefs in Panel B yields consistent, but imprecise, results.

Next, we find that childhood under-confidence in math is starkly associated with major choice among those who go to college (Table 4, columns 5 and 6). Among those with a 4-year college degree, students who were under-confident in math are 18.4pp (se = 4.5pp) less likely to earn a STEM major\textsuperscript{28} than their peers with comparable childhood scores, a 66 percent drop from the share of STEM majors across all college graduates in our sample. This large gap means that under-confident children who score above the 50th percentile on the CDS math test as children are less likely to major in STEM, conditional on going to college, than the average child who scores below the 50th percentile. We obtain very similar results using our more continuous measure of biased beliefs in Panel B: a 1sd increase in confidence is associated with a 9.9pp (se = 2.8pp) increase in the likelihood of majoring in STEM.

Finally, we find no significant relationships between biased beliefs and getting a graduate degree, though again our standard errors are large (Table 4, column 7 and 8).

6.4 Employment outcomes

Next, we examine the links between childhood over- and under-confidence in math and employment outcomes in young adulthood: occupation type, earnings, and employment status. We follow respondents in the adult PSID when they age out of the TAS, so we observe our oldest CDS respondents through age 36 at the end of our sample period. Since respondents’ employment outcomes in their early twenties may not yet be representative of their long-term career trajectories, we restrict the sample to observations in which respondents are older than 25; we observe about 67 percent of our sample above this threshold at least once.

We first consider job choice. Under-confident children are about 7.2pp (se = 2.8pp) less likely\textsuperscript{28} Engineering, math and computer sciences, natural sciences, and health fields.
to work in a STEM occupation\textsuperscript{29} than their peers (Table 5, columns 1 and 2), a gap that is 52% of the baseline rate at which respondents later work in STEM in our sample. We find a similar result with our measure of the degrees of over- and under-confidence, where a 1sd increase in childhood confidence is associated with a 2.1pp (se = 1.2pp) increase in the likelihood that one works in STEM (p-value = 0.09). These confidence gaps are large relative to the links between childhood math scores and later STEM employment, which are generally indistinguishable from zero (Appendix Table A12, column 9).

On the other hand, there are no gaps in the likelihood that over- or under-confident children work in non-STEM high-education occupations\textsuperscript{30} (Table 5, columns 3 and 4). These results are reassuring for our empirical design: the fact that math confidence matters for STEM employment, but not other high-education employment, helps to validate that we properly isolate long-term associations with children’s biased beliefs in math, rather than picking up correlations with unobserved self esteem or other abilities. Taking these point estimates at face value, about one third of the under-confident children who do not pursue STEM careers switch into other high-education occupations, while the rest pursue other work. However, our 95-percent confidence intervals include estimates suggesting that under-confident children are up to 4.0pp less likely or up to 9.0pp more likely to work in other high-education occupations than their peers.

Next, we consider respondents’ earnings. Our regression results are imprecisely estimated, but they broadly suggest that higher math confidence is associated with higher earnings later in life. Using our binary measures of over- and confidence, our point estimates suggest that over-confident children earn about 6 percent (se = 8 percent) more and under-confident children earn about 7 percent (se = 6 percent) less than others with comparable childhood scores, though neither result is statistically significant (Table 5, columns 5 and 6). Using our more continuous measure

\textsuperscript{29}We define STEM fields to include computer and mathematical occupations, architecture and engineering occupations, life, physical, and social science occupations, and healthcare practitioners and technicians. We find similar results when we estimate results for non-healthcare STEM and healthcare occupations separately.

\textsuperscript{30}We define non-STEM high-education occupations as management, business, and financial occupations, legal occupations, education, training, and library occupations, and occupations that focus on writing and communication (a subset of media, arts, and entertainment occupations).
of confidence yields more precise results: a 1sd increase in confidence in childhood is associated with 5.6 percent (se = 2.9 percent) higher earnings in adulthood. This gap is large relative to the association between childhood math scores and adult earnings: increasing test scores by one decile is associated with 8.5 percent higher earnings on average (Appendix Table A12, column 11).

Finally, we consider unemployment. Our regressions suggest that higher confidence may be associated with lower unemployment risk (Table 5, columns 7 and 8). While our binary indicators for over- and under-confidence are not significantly associated with unemployment (Panel A), a 1sd increase in childhood confidence is associated with a 2.3pp (se = 0.9pp) lower likelihood of having been unemployed in the previous year. This gap is large relative to the association between childhood math scores and unemployment: increasing test scores by 10 percentiles is associated with 1.1pp lower unemployment risk on average (Appendix Table A12, column 12).

While most of our results are quite stable – both in magnitude and precision – to the many robustness tests we run in Section 7, these last two results (earnings and unemployment) should be interpreted with caution. They are only statistically significant when using our more continuous measure of biased beliefs, which is more vulnerable to measurement error, and we will show in Section 7 below that they are not robust to using measures of confidence that minimize measurement error by using data from two waves of the CDS. That said, they are suggestive and are consistent with our other findings on the long-term links between childhood confidence and later-life outcomes.31

6.5 Over- versus under-confidence

One ex-ante strength of our binary measures of biased beliefs is that they offer a clear way to test whether over- and under-confidence correlate with later-life outcomes with symmetric magnitudes; we display p-values for all of these comparisons at the bottom of Panel A in Tables 3, 4, and 5. In practice, we find that the coefficient magnitudes for over- and under-confidence are only

31 Given the large gender gaps in confidence we documented in Section 4, we also test whether the relationships between over- and under-confidence and all of our outcomes vary by gender. Our results are primarily driven by boys, though coefficients for boys and girls generally have the same sign and are statistically indistinguishable from each other (available upon request).
significantly different for two of our twelve outcomes: high-school graduation and and working in STEM. Over-confidence predicts high-school graduation significantly more strongly than does under-confidence, while only under-confidence predicts working in STEM.

We also test for heterogeneity in the direction of biased beliefs using our more continuous measure of degrees of confidence. In Appendix Table A13, we allow the coefficient on this measure to differ by whether a child is over-confident (assessing one’s ability at least 3 bins, or 42 percentiles, too high), under-confident (assessing one’s ability at least 3 bins too low), or neither. We cannot reject that the slope of the outcome with respect to the degrees of confidence variable is equal across these groups for any outcome, though we are likely under-powered to do so. This result supports the functional-form assumptions we make in Panel B of each of our main tables, where degrees of confidence enter linearly for all outcomes. More broadly, these results and those using our binary measures of over- and under-confidence suggest that over- and under-confidence largely predict similarly-sized, oppositely-signed gaps in long-term educational and employment outcomes.

7 Robustness

In this section, we show that our main results are robust to controlling for a range of possible confounding variables and to many alternate definitions of our key measures of biased beliefs.

7.1 Key confounders: Personality, adult investment, and school quality

First, we show that math over- and under-confidence predict long-run outcomes beyond their correlation with (1) more commonly-studied personality traits, (2) parent and teacher beliefs and investment, and (3) and elementary school quality at the time at which we observe confidence. We do not control for these variables in our main specifications because they are likely jointly determined with math confidence, but they may be important confounders of the links we estimate.

Section 1 of Appendix Table A14 adds controls for children’s Big Five personality traits: conscientiousness, agreeableness, neuroticism, openness, and extroversion. The CDS did not administer standard psychometric scales to identify the Big Five traits among children, so we construct these measures from caregivers’ reports of child behavior (see Appendices B and D for more details.) These more commonly-studied traits are largely distinct from biased beliefs in our sample,
explaining less than two percent of the variation in our measures of over and under-confidence, but they could confound the long-term associations that we observe: other work shows that Big Five personality traits correlate with contemporaneous educational and employment outcomes (e.g. Almlund et al. 2011; Heckman et al. 2019), and we find some correlations between these common personality traits and measures of over- and under-confidence in our sample (Appendix Table A15). However, controlling for childhood personality traits leaves our estimates of the links between over- and under-confidence and long-run outcomes broadly the same.\(^\text{32}\)

Next, we add controls to our main specification for parent investments, like reading or doing homework with the child, teacher ratings of children’s academic, social, and physical competence, and the educational attainment that parents and teachers predict for the child. Note that we only observe teacher perceptions for 20-34 percent of the sample; see Appendix D for more detail on these measures and data limitations. Parent and teacher beliefs and investments likely shape and are shaped by children’s self-perceptions in math; if these adults’ investments affect children’s later-life success, they may drive the links between childhood math confidence and later outcomes that we observe (Papageorge et al., 2018; Dizon-Ross, 2019). Measures of teacher and parent beliefs and investments do correlate with children’s beliefs in math in our sample, so they could be important channels through which children’s over- and under-confidence relate to long-run outcomes (Appendix Table A16; discussion in Appendix D). However, Section 2 of Appendix Table A14 shows that children’s over- and under-confidence continue to predict long-run outcomes in similar ways when we add controls for adult perceptions and investment to our main regressions.

Finally, we show that our results are robust to controlling for the quality of school a child attended at the time at which we first observe their over- or under-confidence in math. School quality could confound our main results if it both affects children’s later-life outcomes and shapes their self-assessments in math. In particular, children may assess their own ability relative to their peers, not the national distribution; over-confident children could just be those with relatively

\(^{32}\)Appendix Table A20 shows the coefficients on the personality measures in this regression; they correlate with long-run outcomes in expected ways (Almlund et al., 2011).
low-performing peers, for example. However, we would expect these patterns to bias our results *towards zero*, since later-life outcomes may be worse (better) for children from lower-performing (higher-performing) schools. We use restricted data from the CDS on the school that each student attended to match students with data on the percent of students at their school who qualified for free or reduced-price lunch (a proxy for income), the average student-teacher ratio at their school (a proxy for educational inputs), and levels and trends of their school’s mean achievement levels in math and reading. ³³ Reassuringly, our results do not change meaningfully when we control for school quality (Appendix Table A14, Section 3).

**7.2 Identifying biased beliefs**

In this section, we show that our results are robust to (1) changes in how we define over- and under-confidence indicators, (2) changes in how we construct our more continuous measure of biased beliefs, (3) whether we drop children who score in the top and bottom 10 percentiles of the WJ-R math test from our sample, since they are more likely to be mechanically over- or under-confident, and (4) constructing measures of biased beliefs using data from two waves of the CDS where available, rather than just first-observed measures of confidence. None of these changes affects our main conclusions: that over- and under-confidence strongly and meaningfully predict long-term education and working in STEM.

**Redefining over- and under-confidence:** First, we construct a series of alternate measures of over- and under-confidence that rely on different CDS math score and self-report cutoffs, making those designations more or less strict than our main measures. Appendix E describes these alternate cutoffs in detail. Second, we construct more data-driven measures of over- and under-confidence—

³³We collect data on free or reduced-price lunch and student-teacher ratios from the NCES, while we collect data on testing achievement from the Stanford Education Data Archive (SEDA; Fahle et al., 2021). The measures are scaled relative to national grade- and subject-specific test score distributions. SEDA’s data for school test scores pools data from 2009-2018; unfortunately, the data is not available in earlier years. The students in our sample attended these schools in 1997, 2003, or 2007; we are forced to assume that relative school quality was similar in the decade before we observe testing data. 60 percent of our sample attends a school where we observe test scores in 2009-2019; 80 (50) percent of students attend a school where we observe the student-teacher ratio (percent FRPL) in the year in which we observe confidence. We also include an indicator for missing an NCES School ID in the CDS data.
what we refer to as the relative confidence measures—that identify over- and under-confident children as those in the tails of the distribution of math scores at each self-reported ability level. We identify a child as over-confident (under-confident) if they scored in the bottom (top) 25% of children who report the same self-assessed level of math ability; to parallel our main measures, we only identify over-confidence (under-confidence) among those who reported ability below 5 (over 3) on the 7-point scale. Finally, we construct a third class of binary over- and under-confidence measures based on our more continuous confidence measure: we mark a child as over-confident if this measure is greater than 2 and under-confident if it is less than -2. See Appendix E for more detail on how we define these alternate measures.

**Redefining degrees of confidence:** We also test the robustness of our results to how we map self-assessed ability and observed scores to the same scale, the key design choice in our more continuous measure of confidence. In our main measure, we assume that children should have reported the numbered bin from 1 to 7 in which their CDS score falls when test score percentiles are uniformly distributed across 7 bins (i.e. each bin covers about 14 percentiles). We test robustness to two other transformations: the first assumes that children should have reported the bin from 1-7 in which their test score would fall if children had the CDS’ empirical distribution of self-assessments in mind, and the second instead differences the percentile of self-assessed ability and the percentile of demonstrated ability. Again, Appendix E has more detail on these alternate measures. We convert all three measures to standard deviation units to facilitate comparisons.

**Measurement error:** To reduce the likelihood that our results are driven by measurement error, we construct alternate confidence measures using data on self-assessed and demonstrated ability from two waves of the CDS for the 60 percent of children with multiple measures. First, we construct alternate over- and under-confidence measures by averaging children’s test scores and self-reported ability over two waves and then applying our standard cutoff rules to these average scores and self-reports. Next, we construct alternative two-wave measures of confidence gaps by simply averaging children’s two indicators for being over- or under-confident, calculated separately in each wave. We apply the same logic for the more continuous confidence measure: we either
take the average of raw scores and self-assessments across two years, transform those averages to a common scale, and take the difference, or we simply average the continuous measure calculated separately in two years.

**Sample construction:** Regardless of the measure of confidence used, children at the tails of the test score distribution are mechanically the most likely to be identified as having biased beliefs (under-confidence for high-scorers, and over-confidence for low-scorers). Our main specifications include fixed effects for math and reading test score deciles, but we also show that our results are robust to dropping the highest and lowest deciles from the sample.

**Results:** Appendix Figures A4-A16 present specification charts for our main outcomes of interest. They replicate the specifications discussed thus far, plotting regression estimates that iterate over how confidence is defined, whether we use a first-observed or average confidence measure, and the sample included in the regression. For simplicity, Appendix Table A17 also presents a subset of the results from the specification charts: we iterate through alternate definitions of biased beliefs for each outcome, always using the control variables from our preferred specification and the full sample. Panel A shows the results for over- and under-confidence, and Panel B shows the results for our more continuous measure of biased beliefs. Most coefficients that are statistically significant in our main results are remarkably stable, leaving our conclusions unchanged. The only exceptions are our results for earnings and unemployment, which disappear when we use the more continuous measure of biased beliefs based on two waves of the CDS.

8 **Snowballing investment or persistent over- and under-confidence?**

Childhood over- and under-confidence in math are associated with important gaps in educational and employment outcomes down the line, from math performance during adolescence to career choices in young adulthood. In this section, we aim to understand the dynamic patterns through which these confidence gaps open up and persist.

As we outlined in Section 2, over- and under-confidence in math may primarily affect economic outcomes by shaping children’s own investment decisions or those of parents, teachers, or potential employers. However, the timeline of these impacts is unclear. On one hand, the links we estimate
could arise if math confidence produces investment gaps in childhood that in turn snowball through children’s later education and occupational choices, making them more or less likely to graduate from high school, earn a bachelor’s degree, and seek or earn a STEM major or job. On the other hand, if childhood over- and under-confidence in math persist into adulthood, they may have direct effects on choices and performance at each stage of life, conditional on past achievement. For example, persistent over- and under-confidence could affect young adults’ academic ambition, job search behavior, or negotiating tactics, creating gaps in long-run outcomes even between two individuals with the same prior test scores or resumés.

This section explores the following question: can the gaps in later-life outcomes that we observe be fully accounted for by the links between confidence and intermediate educational outcomes that we observe, or could they arise from persistent biased beliefs as respondents age?

8.1 The persistence of childhood confidence in math

While Table 1, discussed in Section 3.2, shows that over- and under-confidence in math persist from childhood into adolescence, we find that childhood biased beliefs persist even until we last observe respondents in the TAS at ages 18 through 27 (Table 6). This persistence is a necessary condition for children’s biased beliefs to have direct behavioral effects on their educational and career choices as they age. This stability in math confidence also provides additional evidence that our measures of over- and under-confidence capture true psychological traits, not just measurement error in the cognitive tests (see Section 3.2).

We use the wealth of questions in the TAS to construct measures of young adults’ confidence in math and reading, generalized academic confidence, career confidence, and general confidence. Unlike our measures of over- and under-confidence from the CDS, these TAS confidence variables are not paired with measures of demonstrated ability in adulthood. However, the ideal regressions would test the links between childhood over- and under-confidence and biases in adult confidence, so as to avoid conflating the persistence of biased beliefs with the links between childhood confidence and adult achievement. We approximate our ideal specification by controlling for adolescent math and reading scores, digit span scores, and general confidence as proxies for adult ability.
Our first outcome is an index for adult confidence in math using the mean of subjects’ standardized responses to the following questions: “How good would you be in a career or job that required you to use math?” and “How good would you be in a career or job that required you to use physical science or technology?” By this metric, childhood math confidence strongly persists into adulthood. Respondents who were over-confident in math as children score about 0.26sd (se = 0.06sd) higher in math confidence as adults than others with comparable childhood test scores, while under-confident children score about 0.25sd (se = 0.05sd) lower (Table 6, columns 1 and 2, Panel A). Likewise, a 1sd increase in our more continuous measure of childhood math confidence predicts 0.17sd (se = 0.02sd) higher math confidence as an adult (Panel B).

We construct a similar measure of adult reading confidence, standardizing subjects’ responses to the following question: “How good would you be in a career or job that required you to read and write a lot?” Here, children who were under-confident in math score about 0.17sd (se = 0.05sd) higher in adult reading confidence than others with comparable childhood test scores (Table 6, columns 3 and 4), despite the fact that childhood biased beliefs in math are largely unrelated to childhood reading attitudes or confidence in our data (results not shown). Our data suggests that this pattern arises because under-confident children are less likely to work in STEM occupations, making them more likely to have a job requiring reading and writing.34

We also construct indices of generalized academic confidence, career confidence, and general confidence in young adulthood. Generalized academic confidence captures respondents’ beliefs in their skill at solving problems, thinking logically, listening, and teaching others; career confidence captures respondents’ belief that they can attain and succeed in their dream job; and adult general confidence captures respondents’ conviction in their ability to lead and supervise, their independence and decisiveness, and their life’s direction (see Appendix B for the variables comprising

34 An alternative explanation comes from the dimensional comparison theory of self-concept (Möller and Marsh, 2013), in which children develop confidence in one subject relative to another and may focus on one skill set as they age. However, we find that childhood math over- and under-confidence do not predict adolescent reading over- and under-confidence. This suggests that math under-confidence predicts a reading-intensive career, in turn predicting higher adult reading confidence, rather than that math under-confidence predicts adolescent reading confidence, in turn predicting working outside of STEM.
Childhood over- and under-confidence in math predict gaps in adult general academic confidence and career confidence (Table 6, columns 5-8). Children who are over-confident in math score about 0.08sd (se = 0.05sd) in adult academic confidence and 0.11sd (se = 0.05sd) higher in adult career confidence than peers with comparable childhood test scores, though the first is only marginally significant. These patterns also appear in our more continuous measure of confidence, where a 1sd increase in childhood math confidence predicts a 0.04sd (se = 0.02) increase in adult academic confidence and a 0.05sd (se = 0.02) increase in adult career confidence. While it is unsurprising that adult math, academic, and career confidence are correlated, it is reassuring that the links between continuous childhood and adult math confidence are 3-4 times as large as those with these other forms of adult confidence.

Finally, there are no significant relationships between childhood math confidence and adult general confidence (Table 6, columns 9-10). We continue to control for childhood and adolescent general confidence here, so these results suggest that math confidence and more general confidence are fairly stable traits: while general confidence correlates with math confidence in childhood, subject-specific confidence in childhood is not significantly linked with the evolution of general confidence as respondents age.\(^\text{35,36}\)

\(^{35}\)To provide additional evidence that our results capture links with math confidence, rather than with general self esteem or ability, we consider a set of placebo outcomes: individuals’ relationship status, general mental health, social anxiety, alcohol consumption, and dangerous behavior as young adults (all from the TAS). We expect each of these outcomes to be affected by general self-esteem, ability, or confidence but not by math over- and under-confidence specifically. Remarkably, we generally find no relationship between childhood math over- and under-confidence and any of these placebo outcomes, except that math over-confidence predicts a lower likelihood of being in a romantic relationship (Appendix Table A21).

\(^{36}\)To better understand the psychological mechanisms that could explain our main findings, we test whether childhood biased beliefs predict changes in adult personality using measures of respondents’ Big Five personality traits in the TAS (Appendix Table A22, where we estimate our main specification plus controls for childhood personality). Higher math confidence in childhood predicts lower adult “agreeableness” and “openness.” Agreeableness is generally negatively correlated with educational attainment and test scores, while openness tends to have a positive correlation with these outcomes (Almlund et al., 2011). Thus, any impacts of confidence on agreeableness could contribute to our main results, while any impacts on openness would tend to push them towards zero.
In sum, we find that childhood over- and under-confidence in math persist through childhood and into young adulthood as confidence gaps across academic domains and in one’s career. If these biased beliefs have direct effects on respondents’ educational or employment success in adulthood, this persistence may be a key factor in the long-term economic associations that we observe.

8.2 Differential intermediate outcomes do not explain results

Despite the persistence of childhood confidence, the links we observe between childhood biased beliefs and later-life outcomes could still be fully explained by gaps in intermediate educational investments. In Figure 2, we explore the contributions of past investment by estimating the marginal relationships between childhood biased beliefs and later-life outcomes, conditional on all intermediate outcomes that we can observe along the chronological chain of education and entry into the labor market. We then compare these results to those from our baseline specification. If childhood biased beliefs continue to predict longer-run outcomes conditional on intermediate outcomes, the remaining unexplained gaps in outcomes may be related to contemporaneous adult confidence. Of course, this type of analysis is imperfect, especially since we cannot control for all intermediate investments.

Figure 2 reproduces our baseline estimates (Tables 3, 4, and 5, even-numbered columns) for math over- and under-confidence in darker blue, while the lighter blue points present our estimates with controls for all outcomes that precede the outcome of interest. In particular, we re-estimate our regressions for educational outcomes through college holding fixed adolescent math and reading test scores, re-examine having a graduate degree and occupation choice holding fixed all previously-observed education outcomes, and re-examine log earnings and unemployment history with controls for all educational outcomes and past occupation choices.

Many of the large confidence gaps we’ve observed in educational and employment outcomes persist when we condition on observable intermediate outcomes. First, controlling for adolescent academic achievement does not change the relationship between childhood over- or under-confidence and any of our educational outcomes. Turning to job choice, under-confidence is half as predictive of working in STEM when we control for all educational outcomes, including whether
respondents majored in STEM. Finally, gaps in respondents’ earnings fall by up to 60 percent when we condition on intermediate outcomes, though our standard errors remain large. The unemployment coefficients are largely unaffected when we add intermediate outcomes as controls.

In combination with the persistence of math confidence that we documented in the prior section, these results suggest that over- and under-confidence may continue to directly affect economic outcomes as respondents age. In particular, more confident students continue to succeed more in school and invest more in math through college than peers with the same test score history; those educational outcomes then accumulate into long-run gaps in labor market outcomes.

9 Conclusion

In this paper, we identify over- and under-confidence in math outside of the lab and among a large sample of children. In doing so, we are the first to show that even children have markedly biased beliefs about their own ability in math, and that these beliefs are distinct from Big Five personality traits and general confidence. Girls are more likely to be under-confident and less likely to be over-confident in math than boys with the same test scores and general confidence, so gender stereotypes about math may shape perceived ability even at young ages.

We then estimate associations between respondents’ childhood over- and under-confidence in math and their educational and employment outcomes up to 22 years later, including comprehensive controls for children’s demonstrated ability and family backgrounds. The associations we estimate are striking. First, childhood confidence shows important links with the evolution of children’s math achievement through childhood: under-confident children perform worse on the CDS math tests five years later, while over-confident children score higher. Childhood confidence in math also significantly predicts key aspects of later education and work trajectories: whether respondents graduate from high-school and college, their college major choice and occupation choices, their earnings, and whether they experience unemployment. We generally cannot reject that math over- and under-confidence predict outcomes in symmetric ways. We do not observe similar associations with long-run outcomes for childhood confidence in reading, a puzzle that we leave for future work.
Our results suggest that biased beliefs about math ability in childhood may predict later-life outcomes both through accumulated differences in educational investments and by continuing to affect economic outcomes as respondents age. First, childhood over- and under-confidence persist into adolescence and adulthood, and thus could continue to alter respondents’ choices as they age. Second, childhood confidence continues to broadly predict later-life outcomes, particularly in education, when we add controls for all observable educational and career investments along the chronological chain of education and labor market entry.

While our results are not causal, they suggest that confidence in math may crucially shape the education we achieve and jobs we get, with effects possibly taking root as early as childhood. Our results provide key early evidence on the importance of math confidence, but they leave substantial room for future exploration. Besides re-examining the associations we estimate for math and reading over- and under-confidence in an experimental or quasi-experimental setting, research should explore the mechanisms by which childhood math confidence affects later-life outcomes. For example, do less confident children perform worse later because they get less encouragement from teachers, or do they simply choose to exert less effort at school? Next, we’ve seen that high-achievers with low confidence are less likely to work in STEM jobs; do they fare worse in job interviews for those positions, or do they simply not apply? Finally, if future research verifies that confidence causally affects later-life outcomes, what interventions can close those gaps?
References


Mattarella-Micke, Andrew, Jill Mateo, Megan N. Kozak, Katherine Foster, and Sian L. Beilock, “Choke or thrive? The relation between salivary cortisol and math performance de-


Note: We plot first-observed math test scores and self-assessments for the 2985 CDS respondents with at least one year of both measurements. We measure respondents’ ability and self-beliefs in math at ages ranging from 8 to 19, though we observe the median child at 11 and more than 90% of children by age 13. Panel A plots the distribution of respondents’ percentile ranks (calculated relative to a nationally-representative norming sample) on a portion of the Woodcock-Johnson Psycho-Educational Battery Revised (WJ-R) testing math reasoning and knowledge. Panel B plots the distribution of children’s responses when asked to answer “How good at math are you?” on a scale of 1 (not at all good) to 7 (very good). Finally, Panel C plots the average math percentile rank within each category from 1 to 7 of children’s self-reported ability in math.
Figure 2: Controlling for intermediate outcomes

Note: This figure plots the coefficient on over- or under-confidence in our baseline specification (2) and the same coefficient when we add controls for mediating factors. When the outcome is high school or college graduation, majoring in STEM, or college quality, we add controls for adolescent math and reading test scores. When the outcome is going to graduate school, we add controls for all previously-observed education outcomes. When the outcome is occupation choice, we add controls for all observed education outcomes: math and reading scores in adolescence, whether the respondent graduated from high school, college, or graduate school, the 75th percentile of the math SAT score distribution of the college he or she attended, and whether he or she majored in STEM. When the outcome is earnings or unemployment, we add controls for all observed educational outcomes and occupational choice.
Table 1: The persistence of math over- and under-confidence

<table>
<thead>
<tr>
<th>Panel</th>
<th>Measure</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.118***</td>
<td>0.115***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.031)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Panel A. Math over-confidence</td>
<td>N</td>
<td>1747</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample mean</td>
<td>0.041</td>
<td></td>
</tr>
<tr>
<td>Panel B. Math under-confidence</td>
<td>N</td>
<td>1747</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample mean</td>
<td>0.063</td>
<td></td>
</tr>
<tr>
<td>Panel C. Math confidence (SD units)</td>
<td>N</td>
<td>1747</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample mean</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Basic controls: ✓ ✓
Added background controls: ✓

Notes: This table regresses adolescent confidence outcomes on the same confidence outcomes in childhood, including our full set of controls. Adolescent confidence is measured five years after the childhood measurement. Each panel presents a separate set of regressions. In Panel A, we regress a binary variable for math over-confidence in adolescence on a binary variable for over-confidence in childhood. In Panel B, we do the same for binary variables of math under-confidence in adolescence and childhood. In Panel C, we regress our more continuous measure of degrees of confidence in adolescence on the same more continuous math confidence measure in childhood. Here, as throughout our analysis, we standardize this variable to have mean 0 and standard deviation 1 to facilitate interpretation. All of these regressions include either the basic or full set of controls that we use throughout our analysis. All controls that are time-variant are observed in the same year as the confidence measures. Basic controls include child gender, race, decile fixed effects for math and reading test percentile scores, digit span test scores, a general confidence index, family taxable income and its square, parent education, quarter-of-birth fixed effects, year-of-birth fixed effects, age at which confidence was measured fixed effects, and state fixed effects. We also include fixed effects for adolescent test score deciles in math and reading. Added background controls are parents’ rating of child health, indicators for receiving government transfers, household structure, parenting practices, parent occupation, and parent mental health and confidence measures. All controls are recoded to zero if missing and we include a missing indicator. Standard errors are clustered by family, and included in parentheses below each estimate. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively.
Table 2: Demographic predictors of over- and under-confidence

<table>
<thead>
<tr>
<th>Demographic Characteristics</th>
<th>Over-confidence</th>
<th>Under-confidence</th>
<th>Confidence (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>-0.027***</td>
<td>0.023**</td>
<td>-0.097***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Black</td>
<td>0.014</td>
<td>0.015</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.038*</td>
<td>0.019</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Asian or Native American</td>
<td>-0.021</td>
<td>0.024</td>
<td>-0.042</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Only child</td>
<td>0.005</td>
<td>0.016</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>First child</td>
<td>0.015</td>
<td>0.038**</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Second child</td>
<td>0.029*</td>
<td>0.004</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Father graduated high school</td>
<td>-0.011</td>
<td>-0.039</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Father has bachelors</td>
<td>0.014</td>
<td>-0.009</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Mother graduated high school</td>
<td>-0.019</td>
<td>-0.010</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Mother has bachelors</td>
<td>-0.007</td>
<td>-0.025</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Father works in STEM</td>
<td>0.004</td>
<td>-0.036</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Mother works in STEM</td>
<td>-0.010</td>
<td>-0.006</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Father works in non-STEM high-educ</td>
<td>0.000</td>
<td>-0.008</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Mother works in non-STEM high-educ</td>
<td>-0.017</td>
<td>0.019</td>
<td>-0.062*</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Family taxable income (thous 2016 USD)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Other Child Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child ever in gifted prog</td>
<td>0.026**</td>
<td>-0.087***</td>
<td>0.146***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Child ever in special ed prog</td>
<td>0.007</td>
<td>-0.008</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Child has repeated grade</td>
<td>-0.016</td>
<td>-0.012</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Parent’s rating of child health</td>
<td>-0.001</td>
<td>-0.013**</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>
Table 2: Demographic predictors of over- and under-confidence (continued)

<table>
<thead>
<tr>
<th>School Quality Measures</th>
<th>Over-confidence</th>
<th>Under-confidence</th>
<th>Confidence (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent FRPL</td>
<td>-0.027</td>
<td>0.058*</td>
<td>-0.074</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Student-teacher ratio</td>
<td>0.000</td>
<td>-0.000</td>
<td>0.002***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Average math and reading achievement</td>
<td>-0.003</td>
<td>0.003</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Difference btwn math and reading achievement</td>
<td>0.010</td>
<td>-0.004</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Cohort slope of average achievement</td>
<td>-0.036</td>
<td>0.079</td>
<td>-0.228</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>Unable to link to NCES id</td>
<td>0.049*</td>
<td>0.013</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Other Child Ability Measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit span score</td>
<td>-0.000</td>
<td>-0.004**</td>
<td>0.009**</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>General confidence</td>
<td>0.038***</td>
<td>-0.054***</td>
<td>0.211***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>0.085</td>
<td>0.121</td>
<td>0.000</td>
</tr>
</tbody>
</table>

| N                | 2985 | 2985 | 2985 |
| R-squared        | 0.21 | 0.21 | 0.57 |

Notes: Each column regresses a measure of childhood biased beliefs in math on child characteristics. In columns 1 and 2, the dependent variables are our main indicators for over-confidence or under-confidence, respectively. In column 3, the dependent variable is a linear measure of biased beliefs that ranges from -6 to 6, where negative values represent under-confidence, *which we have standardized to have mean 0 and standard deviation one in our sample*. All variables are taken from the first year in which we observe children’s confidence in math. Additional controls include fixed effects for math and reading test score deciles, birth year, birth quarter, state, and age at which confidence was measured. The coefficients on the test score deciles are shown in Appendix Table A11. All controls are recoded to be zero if missing, and the regressions include missing indicators for each variable (not shown). All variables are either continuous or binary indicators, except for child race and birth order. The omitted category for race is non-Hispanic whites, and the omitted category for birth order is any birth order higher than two. Standard errors are clustered by family. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively.
Table 3: Childhood math confidence and medium-term educational achievement and attainment

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Adolescent math scores</th>
<th>Adolescent reading scores</th>
<th>High school degree</th>
<th>College degree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Over-confidence</td>
<td>2.637*</td>
<td>2.629*</td>
<td>-0.362</td>
<td>-0.318</td>
</tr>
<tr>
<td></td>
<td>(1.468)</td>
<td>(1.500)</td>
<td>(1.381)</td>
<td>(1.394)</td>
</tr>
<tr>
<td>Under-confidence</td>
<td>-5.705***</td>
<td>-5.778***</td>
<td>0.353</td>
<td>0.278</td>
</tr>
<tr>
<td></td>
<td>(1.482)</td>
<td>(1.486)</td>
<td>(1.439)</td>
<td>(1.445)</td>
</tr>
<tr>
<td>N</td>
<td>1747</td>
<td>1747</td>
<td>1745</td>
<td>1745</td>
</tr>
<tr>
<td>OC = -1*UC? p-value:</td>
<td>0.147</td>
<td>0.143</td>
<td>0.997</td>
<td>0.984</td>
</tr>
</tbody>
</table>

Panel A: Independent variables are binary measures of over- and under-confidence

Panel B: Independent variable is degrees of over- and under-confidence in standard deviation units

| Confidence          | 2.806***                | 2.800***                  | 0.111             | 0.077         | 0.019*     | 0.018*       | 0.032***   | 0.032***     |
|                     | (0.566)                 | (0.569)                   | (0.587)           | (0.580)       | (0.010)    | (0.010)      | (0.011)    | (0.011)      |
| N                   | 1747                    | 1747                      | 1745              | 1745          | 2714       | 2714         | 2725       | 2725         |

Sample mean of dep. var. | 50.808                  | 48.231                    | 0.876             | 0.297         |

Basic controls: ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓
Added background controls: ✓ ✓ ✓ ✓ ✓ ✓ ✓

Notes: This table regresses educational achievement and attainment outcomes on childhood biased beliefs with various controls. Biased beliefs are measured in the earliest observed wave in the CDS with non-missing test scores and self-assessed ability. In Panel A, the outcome is regressed on an indicator for over-confidence, an indicator for under-confidence and our basic set of controls (in odd-numbered columns) and our extended set of controls (in even-numbered columns). The p-value listed tests whether the coefficient on the over-confidence indicator is equal to -1 times the coefficient on the under-confidence indicator. In Panel B, the outcome is regressed on our more continuous measure of biased beliefs, which we have standardized to have mean zero and standard deviation one in our sample, and the same sets of controls. All controls are the same as described in Table 1, minus the controls for adolescent test score deciles. Standard errors are clustered at the family level and included in parentheses below each estimate. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively.
Table 4: Childhood math confidence and college quality, college major choice, and post-college schooling

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>College quality index</th>
<th>College’s 75th pctile math SAT score</th>
<th>STEM major</th>
<th>Graduate degree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Panel A: Independent variables are binary measures of over- and under-confidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over-confidence</td>
<td>0.067</td>
<td>0.038</td>
<td>6.244</td>
<td>2.891</td>
</tr>
<tr>
<td></td>
<td>(0.145)</td>
<td>(0.147)</td>
<td>(12.112)</td>
<td>(11.837)</td>
</tr>
<tr>
<td>Under-confidence</td>
<td>-0.095</td>
<td>-0.112</td>
<td>-9.312</td>
<td>-10.818*</td>
</tr>
<tr>
<td></td>
<td>(0.081)</td>
<td>(0.081)</td>
<td>(5.958)</td>
<td>(5.900)</td>
</tr>
<tr>
<td>N</td>
<td>1107</td>
<td>1107</td>
<td>1117</td>
<td>1117</td>
</tr>
<tr>
<td>OC = -1*UC? p-value:</td>
<td>0.866</td>
<td>0.663</td>
<td>0.819</td>
<td>0.549</td>
</tr>
<tr>
<td>Panel B: Independent variable is degrees of over- and under-confidence in standard deviation units</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>0.044</td>
<td>0.036</td>
<td>4.198</td>
<td>3.496</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.046)</td>
<td>(3.460)</td>
<td>(3.424)</td>
</tr>
<tr>
<td>N</td>
<td>1107</td>
<td>1107</td>
<td>1117</td>
<td>1117</td>
</tr>
<tr>
<td>Sample mean of dep. var.</td>
<td>0.053</td>
<td>594.172</td>
<td>0.277</td>
<td>0.200</td>
</tr>
<tr>
<td>Basic controls:</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Added background controls:</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Notes: This table regresses college outcomes on childhood biased beliefs with various controls. Biased beliefs are measured in the earliest observed wave in the CDS with non-missing test scores and self-assessed ability. In Panel A, the outcome is regressed on an indicator for over-confidence, an indicator for under-confidence and our basic set of controls (in odd-numbered columns) and our extended set of controls (in even-numbered columns). In Panel B, the outcome is regressed on our more continuous measure of biased beliefs, which we have standardized to have mean zero and standard deviation one in our sample, and the same sets of controls. All controls are the same as described in Table 1, minus the controls for adolescent test score deciles. Standard errors are clustered at the family level and included in parentheses below each estimate. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively.
Table 5: Childhood math confidence and employment outcomes

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Works in STEM (1)</th>
<th>Non-STEM high-educ occ. (2)</th>
<th>Ln(Earnings) (3)</th>
<th>Ln(Earnings) (4)</th>
<th>Unemployed this year (5)</th>
<th>Unemployed this year (6)</th>
<th>Unemployed this year (7)</th>
<th>Unemployed this year (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Independent variables are binary measures of over- and under-confidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over-confidence</td>
<td>-0.009</td>
<td>-0.005</td>
<td>-0.019</td>
<td>-0.026</td>
<td>0.045</td>
<td>0.063</td>
<td>-0.034</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.031)</td>
<td>(0.025)</td>
<td>(0.026)</td>
<td>(0.085)</td>
<td>(0.085)</td>
<td>(0.030)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Under-confidence</td>
<td>-0.075***</td>
<td>-0.072**</td>
<td>0.029</td>
<td>0.025</td>
<td>-0.067</td>
<td>-0.072</td>
<td>0.006</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.056)</td>
<td>(0.057)</td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>N</td>
<td>4592</td>
<td>4592</td>
<td>4592</td>
<td>4592</td>
<td>4423</td>
<td>4423</td>
<td>4975</td>
<td>4975</td>
</tr>
<tr>
<td>OC = -1*UC? p-value:</td>
<td>0.043</td>
<td>0.063</td>
<td>0.822</td>
<td>0.974</td>
<td>0.833</td>
<td>0.927</td>
<td>0.437</td>
<td>0.395</td>
</tr>
<tr>
<td>Panel B: Independent variable is degrees of over- and under-confidence in standard deviation units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>0.022*</td>
<td>0.021*</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.049*</td>
<td>0.056*</td>
<td>-0.023**</td>
<td>-0.023**</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.028)</td>
<td>(0.029)</td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>N</td>
<td>4592</td>
<td>4592</td>
<td>4592</td>
<td>4592</td>
<td>4423</td>
<td>4423</td>
<td>4975</td>
<td>4975</td>
</tr>
<tr>
<td>Sample mean of dep. var.</td>
<td>0.139</td>
<td>0.163</td>
<td>10.185</td>
<td>0.167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic controls:</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Added background controls:</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Notes: This table regresses employment outcomes on childhood biased beliefs with various controls. Biased beliefs are measured in the earliest observed wave in the CDS with non-missing test scores and self-assessed ability. In Panel A, the outcome is regressed on an indicator for over-confidence, an indicator for under-confidence and our basic set of controls (in odd-numbered columns) and our extended set of controls (in even-numbered columns). The p-value listed tests whether the coefficient on the over-confidence indicator is equal to -1 times the coefficient on the under-confidence indicator. In Panel B, the outcome is regressed on our more continuous measure of biased beliefs, which we have standardized to have mean zero and standard deviation one in our sample, and the same sets of controls. All controls are the same as described in Table 1, minus the controls for adolescent test score deciles. Basic controls also include year fixed effects, since these outcomes are observed in a panel. Standard errors are clustered at the family level and included in parentheses below each estimate. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively.
Table 6: Childhood math confidence and young adult confidence outcomes

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Math confidence</th>
<th>Reading confidence</th>
<th>Academic confidence</th>
<th>Career confidence</th>
<th>General confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Over-confidence</td>
<td>0.264***</td>
<td>0.259***</td>
<td>0.002</td>
<td>-0.000</td>
<td>0.090*</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.058)</td>
<td>(0.067)</td>
<td>(0.067)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>Under-confidence</td>
<td>-0.250***</td>
<td>-0.246***</td>
<td>0.159***</td>
<td>0.166***</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.047)</td>
<td>(0.053)</td>
<td>(0.054)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>N</td>
<td>6632</td>
<td>6632</td>
<td>6634</td>
<td>6634</td>
<td>8096</td>
</tr>
<tr>
<td>OC = -1*UC? p-value:</td>
<td>0.850</td>
<td>0.863</td>
<td>0.064</td>
<td>0.057</td>
<td>0.362</td>
</tr>
</tbody>
</table>

Panel A: Independent variables are binary measures of over- and under-confidence

- Over-confidence: 0.264*** (0.058), 0.259*** (0.058), 0.002 (0.067), -0.000 (0.067), 0.090* (0.047), 0.084* (0.048), 0.103** (0.050), 0.106** (0.050), 0.056 (0.043), 0.056 (0.043)
- Under-confidence: -0.250*** (0.047), -0.246*** (0.047), 0.159*** (0.053), 0.166*** (0.054), -0.038 (0.033), -0.042 (0.033), -0.058 (0.038), -0.055 (0.038), 0.002 (0.031), 0.002 (0.031)

Sample mean of dep. var.: -0.000

Panel B: Independent variable is degrees of over- and under-confidence in standard deviation units

- Confidence: 0.167*** (0.021), 0.167*** (0.021), -0.046* (0.026), -0.048* (0.026), 0.037** (0.017), 0.038** (0.017), 0.055*** (0.018), 0.054*** (0.018), 0.020 (0.016), 0.021 (0.016)

Sample mean of dep. var.: -0.000

Notes: This table regresses young adult confidence outcomes on childhood biased beliefs with various controls. Biased beliefs are measured in the earliest observed wave in the CDS with non-missing test scores and self-assessed ability. In Panel A, the outcome is regressed on an indicator for over-confidence, an indicator for under-confidence and our basic set of controls (in odd-numbered columns) and our extended set of controls (in even-numbered columns). The p-value listed tests whether the coefficient on the over-confidence indicator is equal to -1 times the coefficient on the under-confidence indicator. In Panel B, the outcome is regressed on our more continuous measure of biased beliefs, which we have standardized to have mean zero and standard deviation one in our sample, and the same sets of controls. All controls are the same as described in Table 1, minus the controls for adolescent test score deciles. Basic controls also include year fixed effects, since the outcomes are observed in a panel. In this table, we also add controls for adolescent test score deciles in math and reading, as well as adolescent general confidence and digit span scores in all specifications. Standard errors are clustered at the family level and included in parentheses below each estimate. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively.