Measurement and Development of Noncognitive Skills in Adolescence: Evidence from Chicago Public Schools and the OneGoal Program

Tim Kautz
Mathematica

Wladimir Zanoni
Inter-American Development Bank

Using administrative data, we develop measures of noncognitive skills and evaluate OneGoal, an intervention designed to help disadvantaged students complete college by teaching them noncognitive skills. We (1) compare the outcomes of participants and nonparticipants with similar characteristics and (2) use a difference-in-differences approach exploiting that OneGoal was introduced into different schools at different times. We estimate that OneGoal increases college enrollment by 10–20 percentage points for males and females and reduces arrest rates by 5 percentage points for males. Through a mediation analysis, we find that improvements in noncognitive skills account for 13%–32% of these effects.

I. Introduction

Many disadvantaged adolescents do not attain postsecondary degrees. For example, data from Chicago Public Schools (CPS) through 2020 suggest that about 83% of entering ninth graders graduate from high school

We acknowledge support from the ECMC Foundation and the Hymen Milgrom Supporting Organization's Successful Pathways from School to Work Initiative. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of any of the funders. The study received institutional review board (IRB) approval from the University of Chicago's IRB (IRB12-2002). We thank the OneGoal staff and especially Kylie Vadnais for assistance in understanding the OneGoal program and data. We benefited from helpful comments from Amanda Agan, Elaine Allensworth, Enghin Atalay, Stephane Bonhomme, Angela Duckworth, Philipp Eisenhauer, Jorge Luis García, Miriam Gensowski, John Eric Humphries, Magne Mogstad, María Rosales-Rueda, Lisa Rosen, Todd Rosencranz, Bradley Setzler, Melissa Tartari, Jake Torcasso, the participants of the Lifecycle Working Group and Family Economics Workshop at the University of Chicago, and two anonymous reviewers. We thank Peter Porcino

Electronically Published May 15, 2024.

Journal of Human Capital, volume 18, number 2, summer 2024.

https://doi.org/10.1086/728087

within 4 years, but only about 27% earn a bachelor's degree (Maloney et al. 2021). Most educational strategies designed to improve such outcomes focus on cognitive skills as measured by achievement test scores. However, such test scores miss noncognitive skills such as persistence, "grit," curiosity, self-control, and sociability (Heckman and Kautz 2012). These skills are powerful predictors of outcomes and remain malleable throughout adolescence, leaving room for interventions during this crucial period (Roberts et al. 2007; Heckman and Kautz 2014; Kautz et al. 2014). In this paper, we demonstrate that noncognitive skills can be measured by use of administrative data that are readily available from school records and then use those measures to study whether a high school intervention can improve educational outcomes by fostering noncognitive skills in disadvantaged youth.

Some recent interventions for disadvantaged adolescents have focused on noncognitive skill development in their curricula (Heller et al. 2017; Yeager 2017). This paper studies OneGoal, a prominent example of this type of intervention. OneGoal attempts to help disadvantaged high school students successfully transition to college.[1] It adopts some traditional approaches to improvement of outcomes, such as helping students to write applications, to select colleges, and to improve their test scores, but it also teaches noncognitive skills such as time management, goal attainment, teamwork, and self-reflection. We use a novel database of linked administrative records to conduct the first rigorous evaluation of OneGoal to estimate its effect on cognitive and noncognitive skills, educational attainment, and criminality.[2]

The main challenge in evaluating OneGoal is accounting for selection bias—the possibility that the outcomes of participants would differ from those of nonparticipants in the absence of the program. We address this selection problem in two ways. Our first approach compares the outcomes of OneGoal participants with the outcomes of otherwise observationally equivalent high school students who did not participate in OneGoal (those who, in particular, have similar demographic characteristics and levels of preprogram cognitive and noncognitive skills). This approach goes beyond similar evaluations that control only for basic demographics or measures of cognitive skills. We show that accounting for noncognitive skills is important, because compared with nonparticipants, OneGoal participants tend to have higher noncognitive skills before the program.

and Elizabeth Schaefer for excellent editorial assistance. We thank Mariana Calderón Cerbón and Ian Morey for invaluable research assistance. We are especially grateful to Steven N. Durlauf, James J. Heckman, Robert LaLonde, and Stephen Raudenbush for their continued support of this research.

[1] See Tough (2012) for a book that describes the OneGoal program and how it attempts to foster noncognitive skills.

[2] OneGoal started in Chicago and has since expanded to multiple cities across the United States. Our analysis is restricted to Chicago.

Our second approach exploits that OneGoal was introduced into different schools at different times. As a result, students from some schools and cohorts did not have access to OneGoal simply because it was not offered in their schools. We use this feature to compare the outcomes of students who had access to OneGoal with the outcomes of those who did not, a strategy similar to comparing treatment and control groups in an experiment that randomly assigns the opportunity to participate in OneGoal. We use whether students attended a school that offered OneGoal as an instrument for participation in OneGoal and adopt a difference-in-differences specification that accounts for stable differences in unobserved characteristics. The second approach complements the first approach because it makes fundamentally different identifying assumptions.

Our first approach involves developing valid proxies for unobserved cognitive and noncognitive skills. To do so, we apply a factor model to administrative records on school-related behaviors. Our measure of cognitive skills is based on students' achievement test scores, whereas our measure of noncognitive skill is based on the component of their grades, credits earned, disciplinary infractions, and absences that is unrelated to cognitive skills.[3] Our procedure accounts for measurement error and removes the cognitive component from the noncognitive measures. We validate this measure of noncognitive skill by examining its predictive power and find that it rivals and often outperforms achievement test scores in predicting arrests, high school graduation, college enrollment, and college graduation.

Accounting for selection, our estimates suggest that OneGoal improves high school outcomes such as grades, days absent, test scores, and credits earned. It also reduces arrests for males by 5 percentage points and increases college enrollment by 10–20 percentage points for both males and females. The results are robust to many specifications.

This paper contributes to knowledge on skill development in adolescence in at least three ways. First, it includes longer-term outcome data and considers a broader set of outcomes than many previous evaluations of interventions for adolescents.[4] Second, it digs deeper than most evaluations of adolescent programs by conducting a mediation analysis that demonstrates the improvements in outcomes are linked to improvements in skills. Third, it builds on a growing body of relatively recent evidence that

[3] See Heckman, Humphries, and Veramendi (2014) and Jackson (2018) for recent papers in economics that use similar measures. See also Borghans et al. (2011), Duckworth, Quinn, and Tsukayama (2012), and West et al. (2016).

[4] See the discussions in Heckman and Kautz (2014) and Heckman, Jagelka, and Kautz (2021), which summarize the nature and efficacy of a wide range of interventions. Some recent evaluations are important exceptions. For example, Millenky et al. (2011), Heller et al. (2017), and Fein, Dastrup, and Burnett (2021) consider a similarly broad set of outcomes and have relatively long follow-up periods.

suggests adolescent interventions can be effective, providing a counterweight to previous evidence suggesting that adolescence may be too late for successful intervention.[5] Supporting the findings in this paper that OneGoal is promising, a recent unpublished report presents results from an evaluation of OneGoal that uses long-term data to compare OneGoal participants with similar nonparticipants and finds effects on high school and college outcomes similar in magnitude to those that we find (Hallberg et al. 2022). Our paper goes beyond that report by constructing separate measures of cognitive and noncognitive skills, examining effects on arrests, and conducting difference-in-differences and mediation analyses.

Moreover, our analysis illustrates broader points about test scores and noncognitive skills. Before entering the program, OneGoal participants tend to have near-average cognitive skills (test scores) but above-average noncognitive skills compared with their peers. If we had not accounted for baseline differences in noncognitive skills, we would not have adequately modeled selection. On the other hand, if we had studied only the effects of OneGoal on achievement test scores, we would have underestimated the total effect of OneGoal because OneGoal improves outcomes in part because it improves students' noncognitive skills. Our mediation analysis—which allows us to explore these mechanisms—suggests that improvements in noncognitive skills account for 13%–32% of the effects of OneGoal on college enrollment and arrests. More generally, these results reveal the limitations of evaluations (or other studies) that rely solely on achievement test scores to measure skills (Decker, Mayer, and Glazerman 2004; Clark et al. 2013). We show how noncognitive skills can be measured by using administrative records, allowing us to disentangle the mechanisms through which OneGoal affected student outcomes, including the important role of noncognitive skills.

The paper proceeds as follows. Section II provides a description of the OneGoal program. Section III describes the data. Section IV outlines and validates our approach to measuring skills. Section V outlines our approach to estimating treatment effects. Section VI describes the sample and characteristics of students. Section VII provides our main analysis, which includes estimates of the treatment effects and a mediation analysis. Finally, section VIII concludes the paper.

[5] See also Heckman and Kautz (2014) and Heckman, Jagelka, and Kautz (2021) for a discussion. Important examples of promising interventions for adolescents and young adults include Career Academies (Kemple and Snipes 2000; Kemple and Willner 2008), the Year-Up program (Roder and Elliot 2011, 2014; Fein, Dastrup, and Burnett 2021), Becoming a Man (Cook et al. 2014; Heller et al. 2017), the National Guard ChalleNGe Program (Bloom, Gardenhire-Crooks, and Mandsager 2009; Millenky, Bloom, and Dillon 2010; Millenky et al. 2011), the WorkAdvance program (Katz et al. 2022), the Stay the Course program (Evans et al. 2020), the Bottom Line college advising program (Barr and Castleman 2021), the CUNY Accelerated Study in Associate Programs (Sommo et al. 2018; Weiss et al. 2019), and the One Million Degrees program (Hallberg et al. 2023).

II. Overview of OneGoal

Because accounting for selection is the main challenge of this paper, in this section we detail how participants ("OneGoal Fellows") are recruited and selected. OneGoal offers its services to students through a daily, in-school course taught by a "Program Director"—an active CPS teacher who has been selected and trained by OneGoal. OneGoal is a nonprofit that is funded by corporate sponsors, foundations, and private donors. Half of the curriculum focuses on improving "college access," that is, helping OneGoal Fellows improve their grades and test scores, teaching them how to write college essays, and discussing college choices. In addition, the program provides a college visit, a financial aid workshop, a college-essay workshop, and an online ACT preparation course. The Program Directors mentor OneGoal Fellows throughout the first year of college to help them navigate their coursework and other challenges.

The other half of the curriculum provides lessons on how to develop specific noncognitive skills and gives OneGoal Fellows an opportunity to apply the lessons to their schoolwork and the college admissions process. For example, one lesson covers how to set goals and create an action plan to accomplish these goals. OneGoal Fellows then apply this lesson by setting a particular academic goal for themselves and assessing whether their plan succeeded. The logic is that practice reinforces skill development and might also improve intermediate outcomes that are useful for college admissions.

OneGoal Fellows are selected by a multistage process. First, OneGoal selects active CPS teachers (the Program Directors) by checking their references, as well as interviewing and observing them in the classroom.[6] Second, students are nominated by teachers or are targeted through informational sessions. Interested students submit an application, which includes two written essays. Qualified applicants are interviewed and are rated on the "five leadership principles" of OneGoal (professionalism, ambition, integrity, resilience, and resourcefulness) and their interest in completing the program.[7] Among other factors, obtaining a higher rating in these principles increases the likelihood of being offered participation in OneGoal. This feature of the selection process suggests that OneGoal might be selecting students who are more motivated (i.e., have higher levels of noncognitive skills) than the average CPS student. In section VI, we show that, in fact, when compared with their peers in CPS, OneGoal Fellows have near-average cognitive skills but above-average noncognitive skills at baseline.

[6] See table A1 in app. A2 (tables A1–A34 are available online) for more details on the teacher recruitment process.

[7] See fig. A1 in app. A2 (figs. A1–A12 are available online) for how students were assessed on the five leadership principles.

III. Data

A strength of this paper is the use of a dataset of linked administrative records containing data from five sources: OneGoal administrative records, CPS, the Chicago Police Department (CPD), the National Student Clearinghouse (NSC), and the American Community Survey (ACS). We use these data to construct histories of each CPS student who was in ninth or tenth grade between 2003 and 2013.[8] Of the 2,376 students accepted into OneGoal, we matched 2,342 (99%) of them with the CPS data.[9] Table 1 summarizes and defines the variables used in this study.

We use detailed administrative data from CPS on grade point averages (GPAs); absences; credits earned; disciplinary infractions; ninth-, tenth-, and eleventh-grade test scores (the Explore, Plan, and ACT tests); high school graduation status; student addresses; school addresses; race; gender; and age.[10]

Measurement of absences is complicated by the introduction of a computerized system in 2007 that reduced the role of human error in tracking absences and caused a sudden change in the distribution of measured absences. We account for this change by using percentile absences, which we calculate separately for each grade and school year.[11]

CPS records disciplinary infractions that take place in a school or at a school-related function. These infractions are divided into six broad categories or "groups" on the basis of the specific behaviors associated with those infractions.[12] Group 3–6 behaviors typically merit suspension and range from "disruptive behavior on a school bus" and "gambling" (group 3 behaviors) to "attempted murder" and "kidnapping" (group 6 behaviors). Because of the limited number of infractions, we sum the categories and consider the total number of annual incidents from groups 3–6 for each student.

Using geocoded versions of student and school addresses, we calculate the distance that each student lives from his or her school. We also use the addresses to identify each student's census block group (neighborhood), on which additional data are collected by the US Census Bureau.[13] We link each student's census block group to neighborhood information from the ACS. These neighborhood characteristics supplement the control variables available in the CPS data.

[8] See app. A3 (the appendix is available online) for a more detailed description of the data and how we standardized the variables over time.

[9] See table A2 in app. A2 for the number of students in each cohort in each school.

[10] We adopt the "standard GPA calculation" as described in the Chicago Public Schools Policy Manual (Chicago Public Schools 2013).

For a description of the Explore, Plan, and ACT tests, see ACT, Inc. (2007, 2013a, 2013b).

See app. A3 for a detailed discussion of how we standardized the listed variables over time. Some of these variables have been collected for longer periods of time.

[11] See app. A3 for a detailed description of how this change affected the measurement of absences and how we address these issues.

[12] The classifications have changed slightly over time. We track these changes in a series of Chicago Board of Education reports from 2002 to 2012 (Chicago Public Schools 2002, 2003, 2004, 2005, 2006a,2006b, 2007, 2008, 2009, 2010, 2011, 2012b).

[13] See app. A3 for details on how the distances were calculated and block groups were assigned.

TABLE 1
Description of Variables

Variable	Description	Source
Explore test	A multiple-choice achievement test administered in the ninth grade that covers English usage/mechanics, English rhetoric, math, reading, and science	CPS
Plan test	A multiple-choice achievement test administered in the tenth grade that covers English usage/mechanics, English rhetoric, pre-algebra/algebra, geometry, reading, and science	CPS
ACT	A multiple-choice achievement test administered in the eleventh grade that covers English, math, reading, and science	CPS
Percentile absences	Percentile ranking of total absences, standardized by grade and year	CPS
GPA	Grade point average, measured on a 4-point scale	CPS
Credits	Total credits earned during a semester	CPS
Discipline	Total number of major disciplinary infractions	CPS
Cohort	First school year in which a student would have been eligible for OneGoal	CPS
Race	Indicator of whether a student is classified as white, black, Hispanic, or other	CPS
High school enrollment status	Whether a student is actively enrolled, left as a nongraduate, transferred, or graduated	CPS
Distance to school	Total number of miles that a student lives from school	CPS
Arrests	Total number of arrests by semester	CPD
Median household income	Median household income in a student's census block group	ACS, CPS
Percentage of single-parent households	Percentage of single-parent households in a student's census block group	ACS, CPS
Employment rate (ages 16–19)	Fraction of residents aged 16–19 that are employed in a student's census block group	ACS, CPS
Enrollment rate (ages 16–19)	Fraction of residents aged 16–19 enrolled in any school in a student's census block group	ACS, CPS
College enrollment	Whether a student is enrolled in college during a particular semester	NSC
College persistence	Number of cumulative semesters enrolled in college	NSC

We also use CPD arrest data that are linked to students in CPS. The arrest database contains all arrest records since 1999 that occurred in Chicago.

Data on postsecondary educational attainment come from the NSC.[14] For each student who completes high school or earns an alternative diploma, we access the student's college enrollment (and graduation) information from the NSC. The data contain information on enrollment periods, type of institution, and graduation status. We measure persistence by the number of semesters that students were enrolled.

IV. Defining and Validating the Skill-Measurement System

Our analysis requires measuring the skills that affect selection into OneGoal, future outcomes, or both. On the basis of OneGoal's recruitment

[14] See app. A3 for a description of how we cleaned the NSC data.

strategy, noncognitive skills likely play a role in determining who participates. In this section, we discuss our approach to measurement of noncognitive skills by use of observed, real-world behavior as captured by school records (grades, credits earned, disciplinary infractions, and absences).

A. *Using Real-World Behaviors to Measure Noncognitive Skills*

Although both cognitive and noncognitive skills can be defined abstractly, we adopt an operational approach by defining skills in terms of how they are measured (Borghans et al. 2008; Almlund et al. 2011). For example, we define cognitive skill as what is captured by achievement tests. Although this definition does not capture all of the possible dimensions of cognitive skill (Ackerman and Heggestad 1997), it is relevant given the prevalence of achievement tests in educational systems (Heckman, Humphries, and Kautz 2014). Similarly, we operationally define noncognitive skill as the component of academic performance that is unrelated to students' performance (i.e., scores) on achievement tests. In the literature, the category of noncognitive skills includes a wide variety of skills, ranging from persistence to agreeableness (Borghans et al. 2008; Almlund et al. 2011). As described further below, our findings and the past literature suggest the resulting measure of noncognitive skill is related to conscientiousness, defined as "the tendency to be organized, responsible, and hardworking" (American Psychological Association 2007).

Our approach to measurement of noncognitive skills by use of real-world behaviors offers some advantages over standard approaches. Psychologists typically elicit personality traits (noncognitive skills) using questionnaires that ask respondents to rate themselves on a numerical scale, such as "On a scale of 1 to 5, how lazy are you?" Economists have argued that it is valid to measure noncognitive skills by use of a broad class of behaviors. If an outcome or behavior depends on a skill, then the behavior is also a valid measure of that skill after adjusting for incentives and other skills (Heckman and Kautz 2012). We measure noncognitive skills by isolating a common factor that drives grades, absences, disciplinary infractions, and credits earned. These measures are valid because they depend on skills beyond raw intellect. For example, earning of course credits requires performance of a series of tasks that reveal skills, such as showing up to class and completing assignments.[15] This logic has been fruitfully applied by Heckman, Humphries, and Veramendi (2014), who measure noncognitive skills by use of adolescent risky behaviors and data from school transcripts. Jackson (2018) applies a similar approach and demonstrates that the teachers who improve measures of students' noncognitive behaviors are often not the same as those who improve students' test scores.

[15] This idea is not new. Ralph Tyler, one of the creators of the original achievement tests, suggested that test scores should be supplemented with a broader class of behaviors (Tyler 1940).

TABLE 2
Predictive Validity (R^2) from Individual Ninth-Grade Measures on Various Outcomes

Outcome	Explore Test (1)	Absences (2)	Credits (3)	GPA (4)	Discipline (5)	All (6)
	Ninth-Grade Measure					
ACT score (grade 11)	.78	.10	.05	.22	.02	.79
GPA (grade 11)	.21	.20	.28	.49	.05	.52
Absences (grade 11)	.09	.35	.12	.22	.03	.39
Arrested within 4 years	.06	.10	.12	.14	.10	.20
Graduate from high school within 5 years	.11	.23	.36	.35	.06	.41
Enroll in college within 6 years	.15	.12	.16	.20	.03	.25
Graduate from college within 10 years	.17	.09	.07	.17	.01	.23

Sources.—CPS, CPD, and NSC administrative data.
Note.—The table shows the predictive power (R^2) from a regression of the outcomes listed in the left-most column on the ninth-grade measures listed across the column headers. “Explore Test” includes the subscores from the reading, English rhetoric, English usage, science, and math subtests of the Explore test. “Absences” indicates the percentile rank of absences in ninth grade. “Credits” includes two separate variables for fall and spring credits accumulated. “GPA” includes the fall and spring GPAs from ninth grade. “Discipline” is a variable for the total number of group 3–6 disciplinary infractions in ninth grade. “All” includes all variables. The time in years is relative to ninth grade. The number of observations ranges from 14,695 to 23,403, depending on the availability of the outcome data.

Some psychologists have argued that it is tautological to use this approach because it uses behavior to predict future behaviors.[16] Heckman and Kautz (2012, 2014) rebut this view by pointing out that any measure of a psychological trait or skill is ultimately derived from a form of behavior. Psychological assessments are no exception, because they require respondents to fill out questionnaires (which is itself a behavior) or report the types of behaviors that they tend to exhibit. To extract a measure of skill from a behavior requires standardizing for other factors that affect the behavior but do not reflect the skills, such as incentives (Borghans et al. 2008; Heckman and Kautz 2012). We account for this possibility in our measurement approach.

In validating our measure of noncognitive skills, we explore the predictive power of the administrative records by estimating the association between administrative records in ninth grade and outcomes measured later.[17] These data are more predictive of life outcomes than what is typically found for self-reported measures.[18] Table 2 shows the predictive validity (R^2) from regressions of each measure on the outcomes. Column 6 shows the R^2 from use of all measures. Test scores are a relatively poor predictor for many outcomes. For example, test scores explain only about 11% of the

[16] See, e.g., the discussion in Pratt and Cullen (2000) and Benda (2005).
[17] This approach is consistent with the standard definition of predictive validity in psychology that focuses on the association between assessments and later measures (American Psychological Association 2007). These associations do not reflect the extent to which these variables can be used to make out-of-sample predictions.
[18] See Almlund et al. (2011) for a review of studies that use self-reported measures.

variation in completing high school, whereas absences and GPA explain about 23% and 35%, respectively. Our measures might be more predictive than self-reported measures because they reflect actual behavior and thereby avoid a problem known as "reference bias," which arises in self-reported questionnaires when respondents rate themselves in comparison with their peers rather than in comparison with the whole population.[19]

B. Factor Model

We use a factor model to (1) reduce the dimensionality of these administrative data (GPA, credits earned, disciplinary infractions, and subscores on achievement tests) and (2) provide a clearer interpretation of our effect estimates.[20] First, by applying standard methods to explore how many latent factors underlie the data, we determined that two factors are sufficient to explain the variation in measures.[21] We then define a measurement system that allows each measure M_j to depend on a factor that represents cognitive skills (θ_{C}) and one that represents noncognitive skills (θ_{N}). As discussed in Heckman and Kautz (2012, 2014), these measures themselves are forms of behavior and could be influenced by incentives or aspects of a person's situation, which we denote as $\mathbf{W}_j$.[22] In our application, for example, we allow attendance to depend on the distance a student lives from school, which proxies for the level of effort students would have to exert to attend classes. We use a linear model to capture the relationship between the measures and latent variables:

$$M_j = \alpha_{\mathrm{C},j}\theta_{\mathrm{C}} + \alpha_{\mathrm{N},j}\theta_{\mathrm{N}} + \boldsymbol{\beta}_j\mathbf{W}_j + \varepsilon_j,$$

where ε_j is the measurement error, and $\alpha_{k,j}$, $k \in \mathrm{C}, \mathrm{N}$ are the "factor loadings" of skill k on measurement j. We assume that $\varepsilon_j \perp\!\!\!\perp (\theta_k, \mathbf{W}_j)$ and $\varepsilon_j \perp\!\!\!\perp \varepsilon_i$ for $j \neq i$.

We set the scale of the latent variables so that for one measure (k), the factor loading on cognitive skill is one ($\alpha_{\mathrm{C},k} = 1$), and for another measure (l), the factor loading on noncognitive skill is one ($\alpha_{\mathrm{N},l} = 1$).[23] Additionally, we restrict the noncognitive skill factor loading for all achievement test score measures to zero. This factor model is identified if there are at least

[19] For a discussion of reference bias and further examples, see Schmitt et al. (2007), Heckman and Kautz (2014), and Lira et al. (2022).

[20] We could have theoretically included arrests as part of our model, but we did not do so because the measure exhibited relatively little variation among tenth graders, so it would have added little signal to our overall measure.

[21] To test this possibility, we conduct a "scree test" by performing a principal component analysis on the full set of ninth-grade measures. We find that the first two eigenvalues are greater than 1 and the third is less than 1. See figs. A3 and A4 in app. A4.2. The Kaiser criterion with Horn's adjustment for sampling error also suggests two factors (Kaiser 1960; Horn 1965).

[22] See Borghans et al. (2008) and Almlund et al. (2011) for summaries of studies showing the importance of accounting for aspects of the situation when measuring traits.

[23] An alternative normalization that would lead to the same variance explained in the outcomes sets each of the factor variances to one.

two measures of cognitive skill and three measures that depend on both cognitive and noncognitive skills (Anderson and Rubin 1956; Heckman, Humphries, and Veramendi 2018; Williams 2020).[24]

Under these normalizations, the cognitive skill factor represents what is measured by achievement tests (after correcting for measurement error), and the noncognitive skill factor represents the underlying dimension that is captured by the other ninth-grade measures but which is not explained by cognitive skill.[25] In this way, any predictive power of noncognitive skills represents the additional gain from using the other measures to supplement achievement tests. This operational definition is particularly interpretable in the context of the US educational system, which relies on achievement test scores to evaluate students. However, it understates the true importance of noncognitive skills because performance on achievement tests depends on noncognitive skills to some degree (Borghans et al. 2011). Alternative normalizations fit the data similarly well but would lead to a different interpretation. For example, we could have assumed that one of the measures—such as absences—depended only on noncognitive skills and allowed the achievement test scores to depend on both cognitive and noncognitive skills, thereby attributing the common correlation between absences and achievement test scores to noncognitive skills. However, our approach helps ensure that the measure of noncognitive skills is distinct from achievement test scores, which is commonly viewed as a measure of cognitive skill.

C. *Validating the Measurement System*

Because measures of noncognitive skill are vital to our study, we validate them by exploring the extent to which they are associated with future outcomes. We adopt a two-step procedure. First, under the assumptions described in section IV.B, we estimate the distribution of factors $F(\boldsymbol{\theta})$ and predict a vector of cognitive and noncognitive factor scores $\hat{\boldsymbol{\theta}}_i$ using the estimated distribution. To account for measurement error in the factor scores, we adopt the "bias-avoidance" method for calculating the factor score (Skrondal and Laake 2001). Second, we estimate the following equation for each outcome k: $Y_{ki} = \boldsymbol{\alpha}_{Yk}\hat{\boldsymbol{\theta}}_i + U_{Yki}$, where U_{Yki} is an error term for person i. To account for estimation error in $\hat{\boldsymbol{\theta}}_i$, we calculate the standard errors by estimating 400 bootstrap samples.

Figure 1 shows the variance explained by cognitive skills, noncognitive skills, and measurement error for (*a*) each of the measures used in the system and (*b*) the outcomes that we analyze.[26] Although both factors are included in the model simultaneously, they are independent, so they

[24] Appendix A4.2 presents an algebraic proof for this specialized case.

[25] We first estimated the model allowing for the factors to be correlated but rejected that the correlation differed from zero ($p = .99$), so we imposed that the factors are independent ($\theta_N \perp\!\!\!\perp \theta_C$) to improve interpretability.

[26] We do not display the variance explained by distance to school because it accounts for a negligible amount of the variance for all measures.

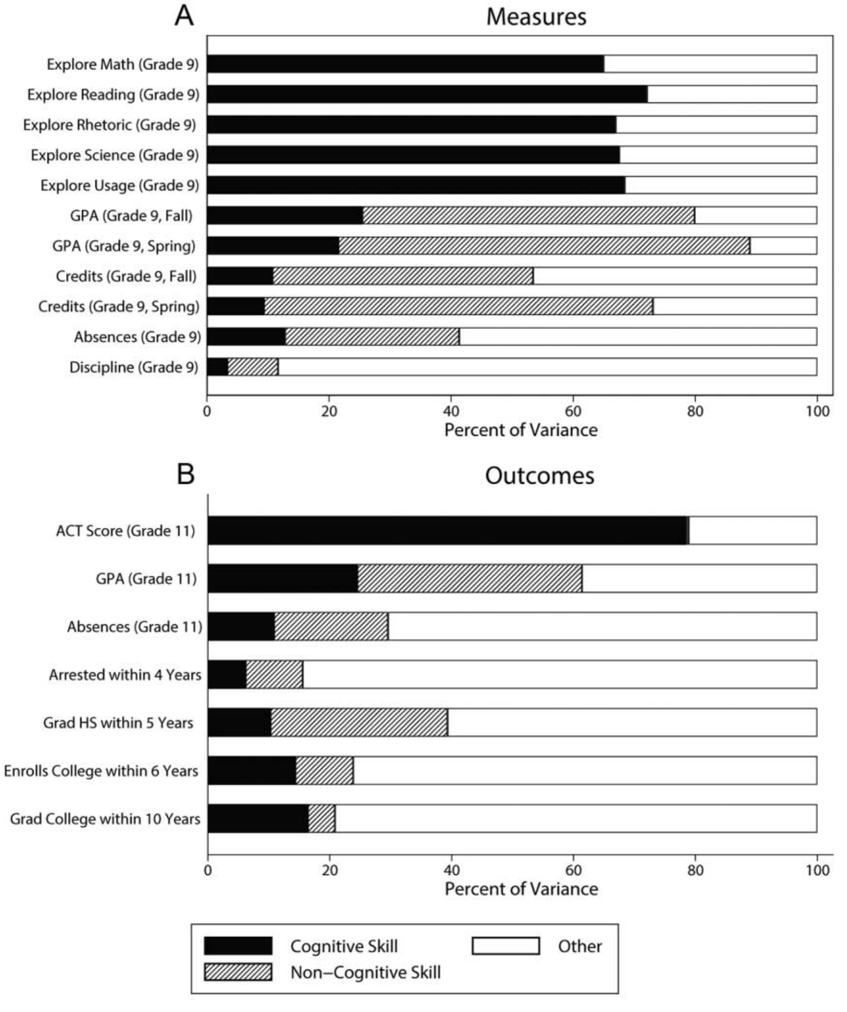

Figure 1.—Variance decomposition of the measurement system and various outcomes. Panel *A* shows a decomposition of the variance of each measure into a component due to cognitive skill, a component due to noncognitive skill, and a residual component assumed to be measurement error. Distance to school explains a negligible amount of the variance for all measures. The subscores on the Explore test are assumed to depend only on cognitive skill. Panel *B* shows a decomposition of the variance of each outcome. The ACT score is not restricted to depend on cognitive skills alone. The number of observations for the measurement system is 23,403. The number of observations from the outcome decomposition ranges from 14,695 to 23,403, depending on the availability of the outcome data. HS = high school. Sources: CPS, CPD, and NSC administrative data.

do not explain common variance in the outcomes. The results mirror those in table 2 and reveal that our measure of noncognitive skill explains much of the variance in meaningful outcomes.

This variance decomposition—coupled with the extant literature—suggests that our measure of noncognitive skill relates to conscientiousness.

Of the administrative measures, the noncognitive skill factor explains the most variance in GPA, suggesting that our measure of noncognitive skill relates to what causes students to earn higher grades that is separate from their performance on achievement tests. Other research has found that measures of noncognitive skills related to conscientiousness are especially associated with GPA, likely because doing well in school requires being organized, responsible, and hardworking (Duckworth, Quinn, and Tsukayama 2012; Borghans et al. 2016), suggesting that our measure also relates to conscientiousness. Past evidence also shows that conscientiousness stands out as an especially strong predictor of educational attainment (Almlund et al. 2011), which aligns with the finding that our measure of noncognitive skill explains a substantial proportion of the variance in high school completion and college enrollment outcomes.

V. Evaluation Methodology

In this section, we describe the approaches we adopt to estimate the effects of OneGoal on outcomes and to conduct a mediation analysis.

A. Approach to Controlling for Preprogram Skills and Characteristics

Our first approach involves estimating the effects of OneGoal by using a model in which we compare OneGoal participants with other students by controlling for unobserved skills and other preprogram characteristics. We adopt a standard potential outcomes framework. For each student, define $A = 1$ if the student applied and was accepted to OneGoal and $A = 0$ if not. Let Y_1 be an outcome if the student were to participate and Y_0 be the student's outcome if not. Let $\mathbf{X}$ be a vector of observed covariates (i.e., basic demographics), and let $\boldsymbol{\theta}$ be a set of unobserved cognitive and noncognitive skills. Our main approach is to proxy the unobserved skills ($\boldsymbol{\theta}$) and compare the outcomes of OneGoal participants with those of other CPS students. This approach relaxes the typical assumption by allowing $(Y_1, Y_0) \perp\!\!\!\perp A|\mathbf{X}$ and instead relies on $(Y_1, Y_0) \perp\!\!\!\perp A|\mathbf{X}, \boldsymbol{\theta}$, where $\perp\!\!\!\perp$ denotes statistical independence.[27]

To implement our analysis, we place several restrictions on the sample. Because charter schools do not report all of the variables we used to measure skills, we restrict the analysis to the non-charter-school sample. Although in principle we could conduct our analysis for the entire sample of schools in CPS, we restrict the sample to schools that at some point offer OneGoal in order to create control groups consisting of students who are most similar to OneGoal participants. We define treatment as whether a student was accepted into OneGoal not whether the student completed the full program to account for the potential for selective attrition out of OneGoal.[28]

[27] See Heckman and Navarro-Lozano (2004) for a discussion of identification.

[28] About 90% of recruits complete the first year of the program. Some students leave OneGoal because they transfer to schools that do not offer OneGoal.

We apply the simple two-step procedure described in section IV.C to estimate the following equation for each outcome k:

$$Y_{ki} = \boldsymbol{\beta}_{Yk}\mathbf{X}_i + \boldsymbol{\alpha}_{Yk}\hat{\boldsymbol{\theta}}_i + \delta_{Yk}A_i + U_{Yki}, \tag{1}$$

where A_i is an indicator for whether person i was accepted into OneGoal (i.e., the person had access to OneGoal and chose to participate, so $D_i = 1$), $\mathbf{X}_i$ are basic demographic characteristics, and $\hat{\boldsymbol{\theta}}_i$ is a measure for the underlying skills. Given OneGoal's focus on postsecondary educational attainment, the main outcomes of interest are whether students enroll in any college and whether they persist in college. Because OneGoal could conceivably affect other outcomes, we also consider more exploratory outcomes, including high school graduation, whether students are arrested, and whether students enroll in a 4-year college. We estimate the equation using ordinary least squares and calculate standard errors by estimating 400 bootstrap samples that allow for errors to be clustered at the school-cohort level—the level at which eligibility for OneGoal varies in our sample. We present the linear regression model as our main approach because it is a transparent version of "parametric matching"(Heckman and Vytlacil 2007), and, as discussed later, the results are similar when using a wide range of alternative methods, including nonparametric approaches and those that do not impose linearity. Because evaluations of past interventions for adolescents have different effects by gender, we conduct the analyses separately for males and females when possible.[29]

B. *Difference-in-Differences Approach*

Our difference-in-differences approach exploits that OneGoal was introduced into different schools at different times, so some cohorts of students did not have access to OneGoal simply because it was not offered in their school. We define access to OneGoal as whether a student was in a school that offered OneGoal when the student was in tenth grade. We use this feature to compare students who had access to OneGoal with those who did not, as in a randomized experiment. To account for both baseline differences across schools and time trends in enrollment rates, we control for school and cohort fixed effects. Ideally, we would explicitly model the time trend for each school before and after the school offered OneGoal, but we have too few pre-OneGoal time periods for many schools. Instead, we assume a common time trend for schools that adopt OneGoal at different times, which is supported by an event-study analysis of the available data. The event study indicates that, although for two of the six outcomes there is some visual indication of pretrends that continue in the post-period, we cannot reject the null that either the preperiod effects are jointly

[29] See, e.g., Kemple and Willner (2008), Rodríguez-Planas (2012), and Carrell and Sacerdote (2013), as well as the discussion in Kautz et al. (2014).

zero or any of the individual preperiod effects is zero (see figs. A7–A12 in app. A7.2). In addition, our approach implicitly assumes that the treatment effects are homogeneous across schools.[30]

We implement the difference-in-differences approach by using access to OneGoal as an instrumental variable. This variable serves as an instrument for participating in OneGoal, because students without access to OneGoal could not have participated, so access is highly correlated with participation by construction. The key identifying assumption is that access to OneGoal is independent of outcomes and the decision to participate. Because students without access to OneGoal could not participate, the instrumental variable estimator consistently estimates the treatment on the treated parameter under very general conditions, including the possibility that students choose to participate in OneGoal on the basis of their perceived benefits of participation.[31]

We use two-stage least squares to estimate the first- and second-stage equations:

$$\begin{aligned} A_{ics} &= \boldsymbol{\beta}^0 \mathbf{X}_{ics} + \delta^0 Z_{ics} + f_c^0 + f_s^0 + \varepsilon_{ics}^0, \\ Y_{ics} &= \boldsymbol{\beta} \mathbf{X}_{ics} + \delta A_{ics} + f_c + f_s + \varepsilon_{ics}, \end{aligned}$$

where c is the cohort, s is the school, and i is the individual, A_{ics} is an indicator variable for whether a student participated in OneGoal, Z_{ics} is an indicator for whether a student had access to OneGoal, f_c is a fixed effect for the cohort, and f_s is a fixed effect for the school. Consistent with our first approach, we allow errors to be clustered at the school-cohort level.

This method yields estimates that are less precise than those of our first method, so we take several steps to increase statistical power. In our first approach, we did not make use of the charter schools that offered OneGoal because they do not report the measures that we use to estimate the distribution of skills. In this section, we use school fixed effects to account for differences between schools, so it is less necessary to control for skills across schools, and therefore we include charter schools in the sample. The estimates are similar when we restrict the sample to exclude charter schools but are less precise because we exclude roughly one-third of the sample. In order to increase statistical power, we combine males and females for this analysis.[32] To limit the loss of power from reduced degrees of freedom, our main specification does not include covariates.

[30] See, e.g., de Chaisemartin and d'Haultfoeuille (2020) for a discussion of this issue. At the time of the publication of this article, we were unaware of peer-reviewed methods to address this issue for nonstaggered designs in which the treatment may switch to being unavailable in one group after it was available, such as in this study.

[31] See Heckman and Vytlacil (2007) for a derivation and discussion in the case of experiments.

[32] The estimates are similar when analyzing males and females separately, but they are estimated less precisely.

C. Mediation Analysis

To investigate the extent to which OneGoal improves later outcomes by improving cognitive and noncognitive skills, we conduct a mediation analysis. We first consider how OneGoal affects cognitive and noncognitive skills in eleventh grade and how changes in skills are associated with improvements in our main outcomes. We then decompose the effect of OneGoal on outcomes into its effect on cognitive skills, noncognitive skills, and other factors. Let $\theta^0_{\mathrm{C}i}$ and $\theta^0_{\mathrm{N}i}$ be cognitive and noncognitive skills in tenth grade, before students are recruited into OneGoal, and let $\theta^1_{\mathrm{C}i}$ and $\theta^1_{\mathrm{N}i}$ be cognitive and noncognitive skills in eleventh grade, the first year in which students are eligible to participate.

Following studies in economics that model skill formation, we allow past skills and covariates to affect future skills.[33] We also allow for the possibility that OneGoal participation could incrementally improve skills in tenth grade:

$$\begin{aligned}\theta^1_{\mathrm{C}i} &= \gamma_{\mathrm{C}0} + \gamma_{\mathrm{C}1}\theta^0_{\mathrm{C}i} + \gamma_{\mathrm{C}2}\theta^0_{\mathrm{N}i} + \phi_{\mathrm{C}}A_i + \boldsymbol{\gamma}_{\mathrm{C}3}\mathbf{X}_i + \eta_{\mathrm{C}i},\\ \theta^1_{\mathrm{N}i} &= \gamma_{\mathrm{N}0} + \gamma_{\mathrm{N}1}\theta^0_{\mathrm{C}i} + \gamma_{\mathrm{N}2}\theta^0_{\mathrm{N}i} + \phi_{\mathrm{N}}A_i + \boldsymbol{\gamma}_{\mathrm{N}3}\mathbf{X}_i + \eta_{\mathrm{N}i},\end{aligned}$$

where $\eta_{\mathrm{C}i} \perp\!\!\!\perp \eta_{\mathrm{N}i}$. We allow the final outcomes Y_{ki} to be a function of eleventh-grade skills, OneGoal participation (A_i), and other covariates:

$$Y_{ki} = \boldsymbol{\beta}_{Yk}\mathbf{X}_i + \boldsymbol{\alpha}_{Yk}\boldsymbol{\theta}^1_i + \delta_{Yk}A_i + U_{Yki}.$$

The total effect of OneGoal is decomposed as follows:

$$\text{Total effect} = \underbrace{\boldsymbol{\alpha}_{Yk}\,\boldsymbol{\phi}}_{\text{Indirect effect}} + \underbrace{\delta_{Yk},}_{\substack{\text{Effect through other}\\ \text{skills or information}}}$$

where $\boldsymbol{\phi} = [\phi_{\mathrm{C}}, \phi_{\mathrm{N}}]$. We estimate the model using a two-stage maximum likelihood approach and calculate the standard errors using 400 bootstrap draws (Heckman and Pinto 2015).

VI. Description of the Final Sample

In this section, we discuss how we construct our sample of students and their characteristics at baseline. We restrict the sample to exclude "selective enrollment" schools because OneGoal does not offer its services to these schools. In order to make the comparison groups as similar as possible, we also restrict the sample to students who were in a CPS school during the second semester of tenth grade.

There are three features of our data that affect our analyses. First, not all schools are required to report all academic indicators. Charter

[33] See, e.g., Cunha and Heckman (2007, 2008) and Cunha, Heckman, and Schennach (2010).

schools do not report absences, grades, or disciplinary infractions. Of the 34 schools that OneGoal has served within Chicago, 12 are charter schools. In section VII.A, we show that controlling for these academic measures is important; thus, most of our analysis is confined to non-charter schools. Our final sample consists of 2,347 OneGoal participants, 59,306 nonparticipants from OneGoal schools, and 186,707 nonparticipants from other schools. For about two-thirds of the sample, we have data on their academic measures in tenth grade.

Second, OneGoal was introduced into different schools at different times. OneGoal began in 2007 in three schools and gradually expanded. This expansion has pros and cons for this evaluation. On the one hand, it limits the number of OneGoal participants who have had a chance to attend and complete meaningful amounts of college. On the other hand, it provides natural control groups in the form of students in OneGoal schools before OneGoal was introduced. Third, we do not observe data from CPS or NSC for students who transfer out of CPS during high school, so we treat these data as missing once students transfer.[34]

We distinguish between OneGoal participants ("participants"), nonparticipants who attended a school that at some point offered OneGoal ("OneGoal school nonparticipants"), and nonparticipants who attended a school that never offered OneGoal ("non-OneGoal-school nonparticipants"). We separate these groups here because we conduct separate analyses using comparison students for OneGoal schools and non-OneGoal schools to rule out the possibility that spillover effects within schools drive the results.

Figure 2 displays the characteristics of OneGoal participants, nonparticipants in OneGoal schools, and nonparticipants from non-OneGoal schools, sorted by gender. Most of the differences between OneGoal participants and other students are statistically significant. However, the magnitude of the difference in test scores is small. Compared with nonparticipants in OneGoal schools, participants score between 0.4 and 0.9 points better on the Plan achievement test, a tenth-grade achievement test designed to be similar to the ACT. At the average score in CPS, a 1 point difference translates to roughly a 10 percentile difference in the national distribution (ACT, Inc., 2013b). By this measure, OneGoal students are near average within CPS.

However, on a range of other measures, OneGoal students are very different from other CPS students. Male participants are about 9 percentage points less likely to be arrested during tenth grade compared with nonparticipants. Both male and female participants have higher GPAs, complete more credits, and have far fewer absences.

[34] Although the NSC does have nationwide coverage, CPS does not request NSC data on students who have transferred out of the district, so we were not able to obtain these data for our sample.

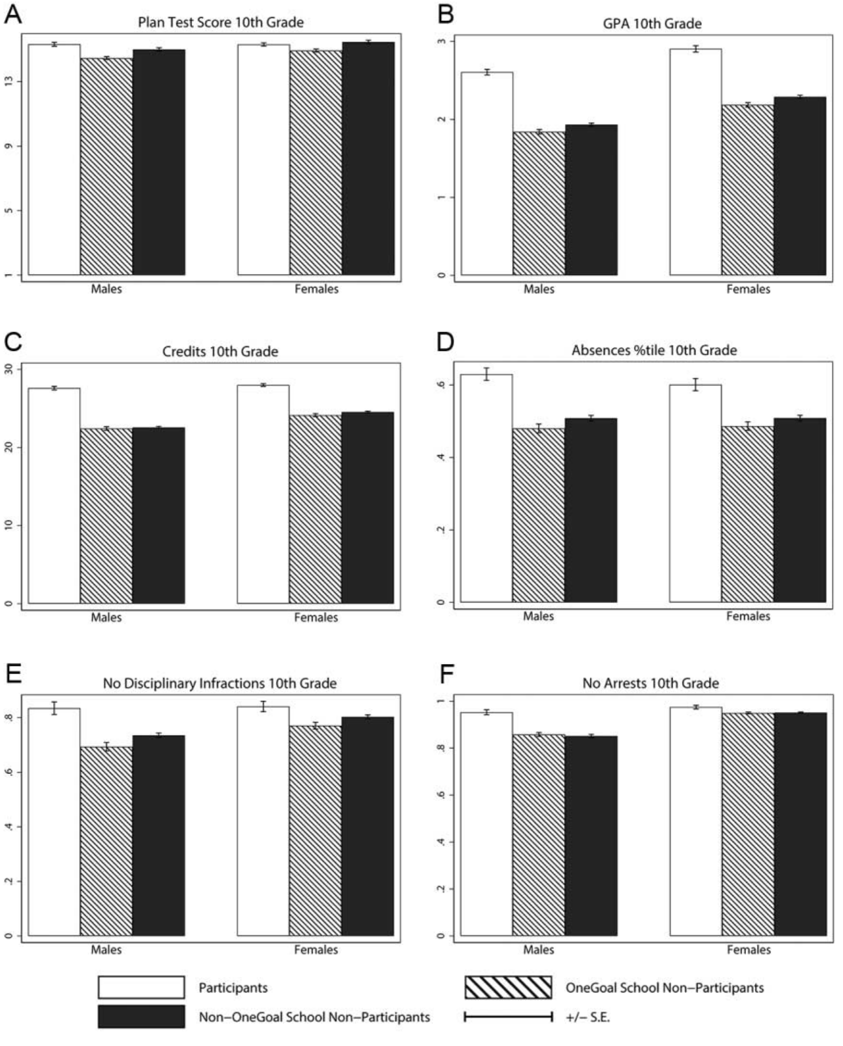

Figure 2.—Preprogram characteristics of OneGoal participants and nonparticipants (tenth grade). The graphs show the average tenth-grade measures for OneGoal participants, nonparticipants in OneGoal schools, and nonparticipants in other schools. The whiskers represent the standard errors for each estimate. All variables have been normalized so that higher values represent beneficial outcomes. "Plan Test Score 10th Grade" is the composite score from the first attempt on the Plan test. "GPA 10th Grade" is the grade point average from the fall and spring semesters of tenth grade. "Credits 10th Grade" is the average credits per semester in tenth grade. "Absences %tile 10th Grade" indicates the percentile rank of absences in tenth grade. "No Disciplinary Infractions 10th Grade" is an indicator for whether a student did not have any group 3–6 disciplinary infractions in tenth grade. "No Arrests 10th Grade" is an indicator for whether a student was not arrested during tenth grade. The standard errors allow for clustering at the school-cohort level. For males, the number of observations ranges from 422 to 896 for the OneGoal participants, 16,755 to 20,295 for the nonparticipants in OneGoal schools, and 45,817 to 55,538 for the nonparticipants in non-OneGoal schools. For females, the number of observations ranges from 509 to 1,005 for the OneGoal participants, 17,073 to 20,566 for the nonparticipants in OneGoal schools, and 47,407 to 57,865 for the nonparticipants in non-OneGoal schools. Sources: OneGoal, CPS, and CPD administrative data.

These patterns suggest that OneGoal participants have higher noncognitive skills than their peers in CPS. We summarize these differences by applying the factor structure described in section IV.B. Figure 3 shows the distribution of extracted cognitive and noncognitive skill factor scores for OneGoal participants, nonparticipants from OneGoal schools, and nonparticipants from other schools. The scores are standardized to have a mean of 0 and a standard deviation of 1 for the full CPS sample. The distribution of cognitive skills is similar for OneGoal participants and nonparticipants, suggesting that cognitive skill plays little role in the selection

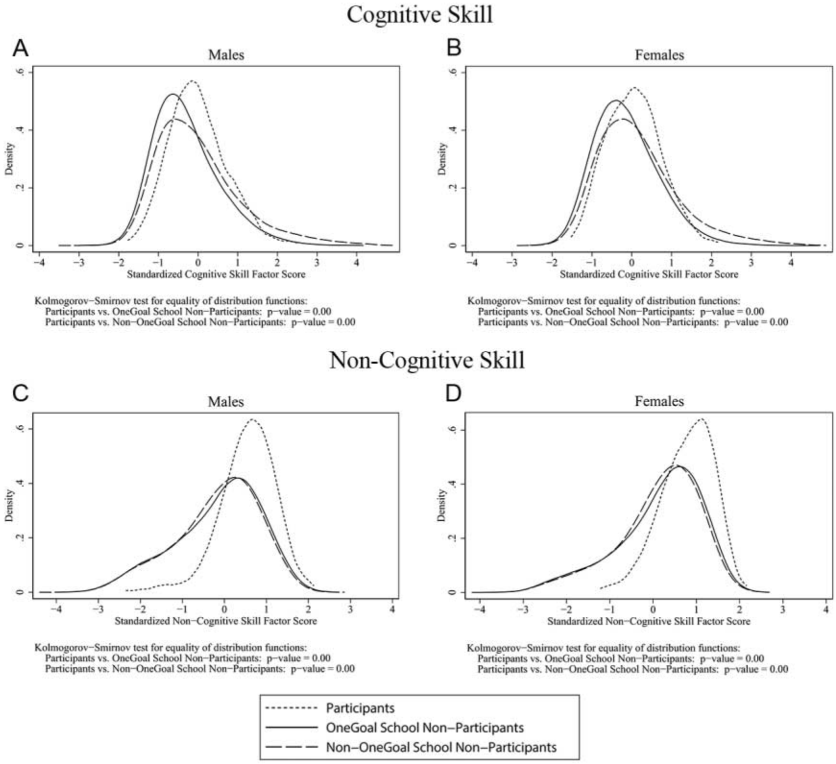

Figure 3.—Distribution of cognitive and noncognitive skills for OneGoal participants and nonparticipants. The top panels show the distribution of predicted cognitive skill factor scores, which are based on the subscores from the reading, English rhetoric, English usage, science, algebra, and geometry subtests of the Plan test. The bottom panels show the distribution of predicted noncognitive skill factor scores, which are based on the fall and spring GPAs from tenth grade, percentile rank of absences in tenth grade, credits accumulated in the fall and spring of tenth grade, and total group 3–6 disciplinary infractions in tenth grade. The noncognitive measures are also allowed to depend on the cognitive measures. The scores have been standardized to have a mean of 0 and a standard deviation of 1 by using a sample of students who were first-time tenth graders in CPS between 2005 and 2013. For males, the number of observations is 512 for the OneGoal participants, 14,879 for the nonparticipants in OneGoal schools, and 41,821 for the nonparticipants in non-OneGoal schools. For females, the number of observations is 587 for the OneGoal participants, 15,400 for the nonparticipants in OneGoal schools, and 46,146 for the nonparticipants in non-OneGoal schools. Sources: OneGoal and CPS administrative data.

process. In contrast, the distribution of noncognitive skills for OneGoal participants is narrower and shifted far to the right compared with those of nonparticipants, suggesting that OneGoal selects students with higher noncognitive skills.[35] Accounting for these preprogram differences is vital for estimating the treatment effects.

VII. Effect Estimates

A. *Estimated Treatment Effects from Controlling for Preprogram Skills and Characteristics*

In this section, we present results from our main approach and summarize a variety of sensitivity analyses. Figure 4 displays results based on the linear model that controls for latent cognitive and noncognitive skills for males and females.[36] For each gender, the first bar shows the difference after controlling only for basic demographics between OneGoal participants and nonparticipants, the second bar shows the effect after additionally controlling for cognitive skills, and the third bar shows the results after additionally controlling for noncognitive skills. The whiskers represent the standard errors for each estimate, and the symbols on the bars indicate the results from tests of significance. Years are measured relative to when students were first in eleventh grade.

The figure reveals three striking results. First, the estimates suggest that OneGoal has positive effects on college outcomes across the board. The effect is the biggest on 4-year college enrollment. Second, OneGoal has greater effects for males than for females. We estimate that OneGoal improves arrest rates for males but not for females, and it has a stronger effect on college outcomes for males. Third, accounting for noncognitive skills is important. If we controlled only for demographics and cognitive skill, we would have estimated that OneGoal increases high school graduation by 10–15 percentage points for both males and females. After controlling for noncognitive skills, we estimate no effect on high school graduation, suggesting that OneGoal recruits the type of students who would have graduated from high school even without the program. This finding indirectly shows the power of noncognitive skills.

We subject these analyses to a series of sensitivity tests that demonstrate the results are robust to many different specifications. In particular, we conduct the following sensitivity checks:

1. Controlling for raw measures of administrative records. In this approach, we include as controls the raw administrative records (achievement test scores, absences, grades, credits earned, and disciplinary infractions) rather than the factor scores based on these

[35] These trends are also apparent when considering the distribution of the individual measures. See tables A14 and A15 in app. A6.

[36] See table A16 in app. A7.1 for the corresponding point estimates.

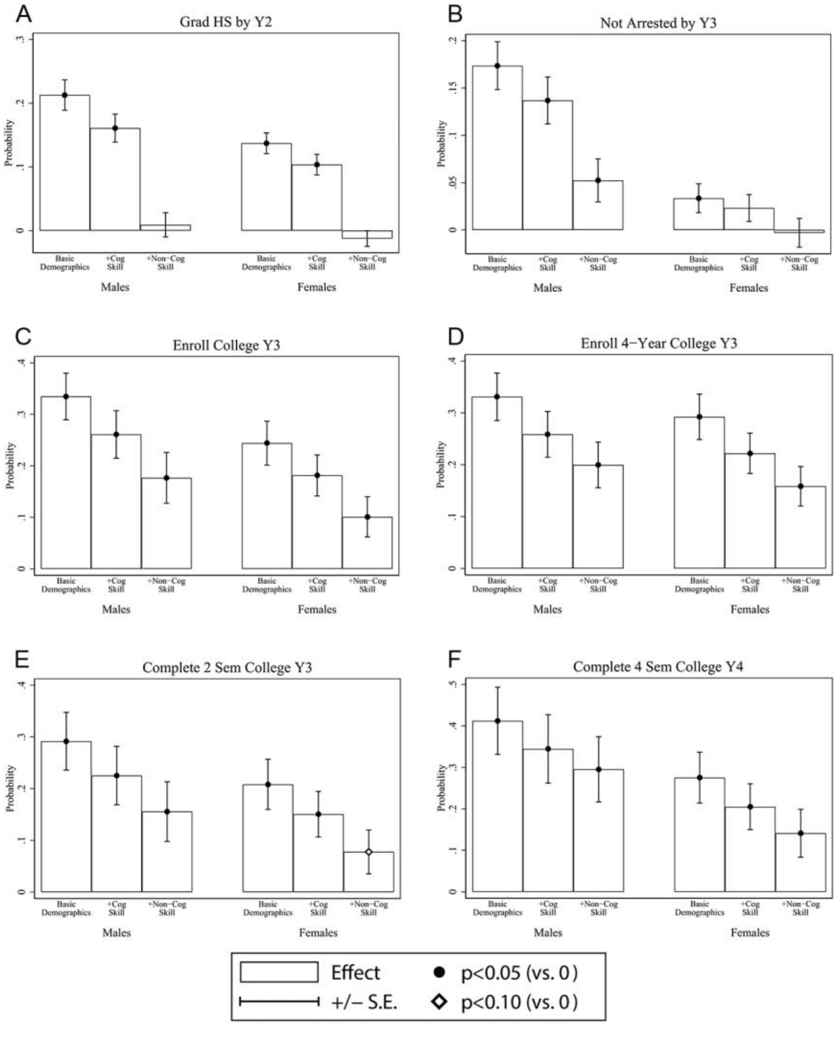

Figure 4.—Treatment effects for main outcomes when controlling for preprogram skills and characteristics. Shown are the effects of OneGoal for each outcome listed at the top of each panel. The labels along the *x*-axis indicate the control variables included. The whiskers represent the standard errors for each mean, and the symbols on the bars indicate the results from tests of significance. "Basic Demographics" include race, cohort, and neighborhood characteristics (median household income, fraction of single-parent households, employment rate, and enrollment rate). "+Cog Skill" includes the basic demographic variables plus a latent cognitive skill factor based on the subscores from the reading, English rhetoric, English usage, science, algebra, and geometry subtests of the Plan test. "+Non-Cog Skill" refers to basic demographics and cognitive skill plus a latent noncognitive skill factor based on the fall and spring GPAs from tenth grade, percentile rank of absences in tenth grade, credits accumulated in the fall and spring of tenth grade, and total group 3–6 disciplinary infractions in tenth grade. The noncognitive measures are also allowed to depend on the cognitive measures. The standard errors were calculated by using 400 bootstrap samples and allow for clustering at the school-cohort level. For males, the number of observations ranges from 9,343 to 13,528, depending on the availability of the outcome data. For females, the number of observations ranges from 10,564 to 14,766, depending on the availability of the outcome data. Sources: ACS and OneGoal, CPS, CPD, and NSC administrative data.

measures. One possibility is that the factor-score approach reduces the dimension of the data at the cost of losing some informational content unique to individual measures. However, we find nearly identical results when using the raw administrative records (for more details, see the discussion and table A18 in app. A7.1).

2. Including schools that never offer OneGoal. Our main analysis restricts the sample of schools to those that offered OneGoal at some point during the analysis period. We find similar results when estimating equation (1) using the full sample of non-charter schools (for more details, see the discussion and table A19 in app. A7.1).
3. Clustering at the school level. For our main specification, we account for clustering at the school-cohort level because treatment status varies at that level (Abadie et al. 2022). For example, some schools offer OneGoal to a cohort of students but not to a subsequent cohort of students. It is more conservative to account for clustering at the school level. However, the results are very similar when doing so (for more details, see the discussion and table A20 in app. A7.1).
4. Including school fixed effects. To avoid reducing the degrees of freedom unnecessarily, our main specification does not include school fixed effects. We reestimate equation (1) including school fixed effects and find very similar results (for more details, see the discussion and table A21 in app. A7.1).
5. Parametric approach allowing for nonlinearities. We also estimate the effects using a two-stage maximum likelihood approach that allows for nonlinearities in outcome variables (Heckman, Humphries, and Veramendi 2014). This approach yields similar results (for more details, see the discussion and table A22 in app. A7.1).
6. Nonparametric approach. To test the robustness of the linear model, we additionally apply a nonparametric approach based on inverse probability weighting. With this approach, we find similar effects of OneGoal as our main analysis (for more details, see the discussion and table A24 in app. A7.1). We conduct a balancing test that demonstrates the baseline covariates are balanced with this approach (see table A23 in app. A7.1).
7. Accounting for students who transferred out of CPS. As noted in section III, we lacked CPS and NSC data for students who transferred out of CPS. In our main analysis sample, 5% of male and 2% of female OneGoal participants transferred out of CPS, whereas 15% of male and 12% of female nonparticipants transferred out of CPS. We estimated that after controlling for demographics and cognitive and noncognitive skills, OneGoal reduced the probability of transferring out of CPS by 3 percentage points for both males and females, although the effect was only statistically significant for females (see table A25 in app. A7.1). Because we found some evidence that OneGoal affected transferring out of CPS, we conducted an analysis

using weights that adjusted for the probability of transferring out of CPS on the basis of students' baseline characteristics. This analysis produced effect estimates that were very similar to our main effect analysis, suggesting that—under the assumption that the observable baseline characteristics accounted for differences in the probability of transferring—missing data from transfer students did not bias our findings (see tables A26 and A27 in app. A7.1).

8. Longer-term follow-up without adjustments. As discussed earlier, our sample is limited in part because charter schools are not required to report all of the measures that we use to proxy cognitive and noncognitive skills. For this reason, our analysis is limited to the first 2 years of college because the sample sizes are too small to conduct the analyses for longer periods. When we include charter schools, we have a larger sample size that allows us to consider college enrollment through the third year of college, but we cannot control for preprogram skills. However, the unadjusted differences are roughly constant over time, suggesting that the effects presented likely persist at least through the third year of college (see fig. A5 in app. A7.1).
9. Controlling for scores from OneGoal application assessments. While our main approach goes beyond most by including measures of noncognitive skills, it depends on having measures of noncognitive skills that adequately capture the selection process. For a set of applicants, we have data on assessments used to evaluate OneGoal applicants on the "five leadership principles" that OneGoal uses for selection (ambition, integrity, professionalism, resilience, and resourcefulness). When controlling for these assessments, the estimates are similar to when controlling for our main set of covariates (see table A31 in app. A7.2). This analysis also accounts for the possibility that the students who choose to apply to OneGoal might be more motivated because it includes those that applied and were not accepted.

B. Estimated Treatment Effects from the Difference-in-Differences Approach

We complement our first analysis with a difference-in-differences approach in which we use access to OneGoal as an instrument for participation. Access is defined as whether a student was in a school that offered OneGoal when the student was in tenth grade. Figure 5 shows the empirical results from this analysis. As with the first analysis, we find that OneGoal has a strong effect on college outcomes but no statistically significant effect on high school graduation. The F-statistic from the first stage is over 50 for each outcome, suggesting that accessibility of OneGoal is not a weak instrument. In a sensitivity check, we include the same demographic covariates as in our first analysis to account for possible changes within

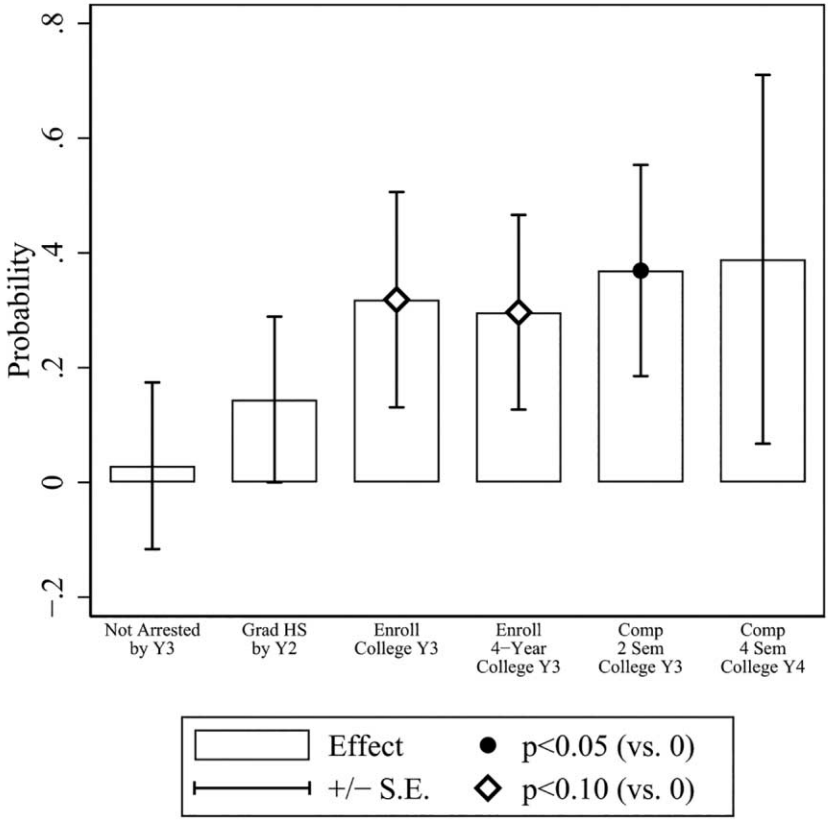

Figure 5.—Treatment effects for main outcomes when using OneGoal eligibility as an instrument. Shown are the effects of OneGoal for each outcome listed along the *x*-axis. The whiskers represent the standard errors for each mean, and the symbols on the bars indicate the results from tests of significance. The standard errors allow for clustering at the school-cohort level. The number of observations ranges from 28,028 to 40,638, depending on the availability of the outcome data. Sources: OneGoal, CPS, CPD, and NSC administrative data.

schools over time. We find similar estimates regardless of whether we include the other covariates, suggesting that school-specific trends do not play a role in these outcomes (see table A32 in app. A7.2). Although they are less precise, the results of this analysis are broadly consistent with our first approach, which is reassuring because the two approaches use very different sources of variation.

C. Mediation Analysis

This section presents a mediation analysis that explores the extent to which the treatment effect on outcomes is due to OneGoal's effects on skills versus other factors. During the first year of the program in eleventh grade, we observe an analogous set of measures to the ones used to estimate

TABLE 3
Treatment Effects for Eleventh-Grade Academic Indicators after Adjusting for Basic Demographics, Cognitive Skill, and Noncognitive Skill

Outcome	Males (1)	Females (2)
ACT score	.50***	.04
	(.14)	(.11)
Absences percentile	.05***	.04*
	(.02)	(.02)
Discipline	.11**	−.05
	(.05)	(.06)
GPA	.13***	.14***
	(.04)	(.04)
Credits	1.32***	.71
	(.47)	(.53)
Minimum observations	12,928	14,836
Maximum observations	16,433	17,450

Sources.—ACS and OneGoal and CPS administrative data.
Note.—The standard errors are displayed below the estimates in parentheses. The table shows the effects of OneGoal for each outcome listed in the left-most column after adjusting for basic demographics, cognitive skill, and noncognitive skill. The basic demographics include race, cohort, and neighborhood characteristics (median household income, fraction of single-parent households, employment rate, and enrollment rate). Cognitive skill is a latent cognitive skill factor based on the subscores from the reading, English rhetoric, English usage, science, algebra, and geometry subtests of the Plan test. Noncognitive skill is a latent noncognitive skill factor based on the fall and spring GPAs from tenth grade, percentile rank of absences in tenth grade, credits accumulated in the fall and spring of tenth grade, and total group 3–6 disciplinary infractions in tenth grade. The noncognitive measures are also allowed to depend on the cognitive measures. The signs of the coefficients for absences and disciplinary infractions have been reversed so that positive values represent beneficial outcomes. The standard errors were calculated by using 400 bootstrap samples and allow for clustering at the school-cohort level.
* Statistically significant at the 10% level.
** Statistically significant at the 5% level.
*** Statistically significant at the 1% level.

the preprogram cognitive and noncognitive skills in tenth grade.[37] We study how the program affects these measures and then place them in the factor framework to estimate the effect on cognitive and noncognitive skills.

Table 3 shows the effects of OneGoal on eleventh-grade academic indicators on the basis of the linear model described in section VII.A. The estimates are adjusted for basic demographics, cognitive skill, and noncognitive skill. The signs of the coefficients for absences and disciplinary infractions have been reversed so that positive values represent beneficial outcomes. For males, OneGoal has a significant effect on ACT scores, absences, credits earned, disciplinary infraction, and GPAs. For females, OneGoal does not have an effect on ACT scores but does improve absences and GPAs. These

[37] We do not have access to a CPS measure of cognitive skill in twelfth grade, so we focus on eleventh grade.

findings suggest that OneGoal might work in part because it improves both cognitive and noncognitive skills.

Figure 6 presents the treatment effects of OneGoal on eleventh-grade cognitive and noncognitive skills for males and females. The measures of skill are standardized by gender to have a standard deviation of 1 for males and females. The findings in this figure are consistent with the patterns observed in table 3. OneGoal improves both cognitive and noncognitive skills in similar amounts for males but improves only noncognitive skills for females.

Figure 7 shows the percentage of the total effect that can be attributed to improvements in cognitive skill, improvements in noncognitive skill, or other factors. We display only the outcomes for which we estimate a statistically significant effect in the analysis presented in figure 4. For males,

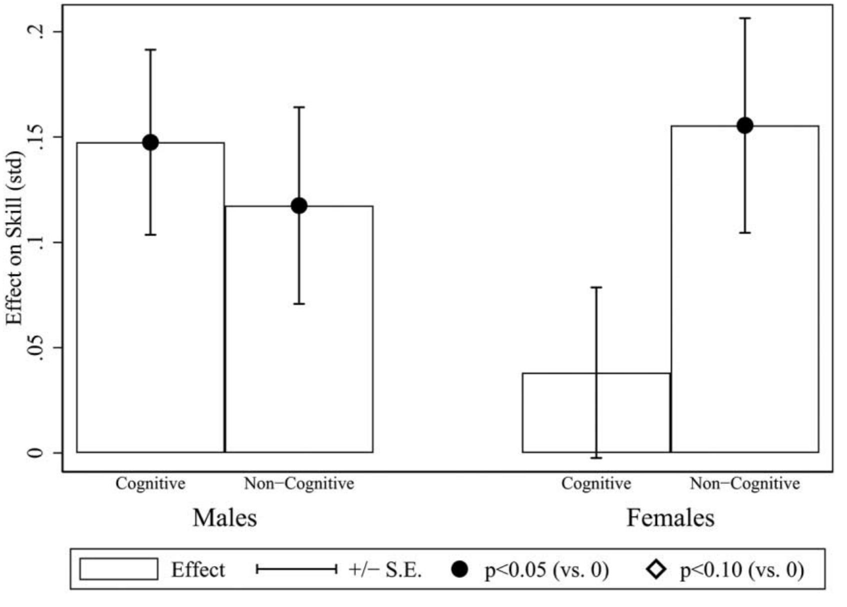

Figure 6.—Treatment effects for eleventh-grade cognitive and noncognitive skills. Shown is the effect of OneGoal on the indicated skill after adjusting for basic demographics, cognitive skill, and noncognitive skill. The basic demographics include race, cohort, and neighborhood characteristics (median household income, fraction of single-parent households, employment rate, and enrollment rate). Cognitive skill is a latent cognitive skill factor based on the subscores from the reading, English rhetoric, English usage, science, algebra, and geometry subtests of the Plan test. Noncognitive skill is a latent noncognitive skill factor based on the fall and spring GPAs from tenth grade, percentile rank of absences in tenth grade, credits accumulated in the fall and spring of tenth grade, and total group 3–6 disciplinary infractions in tenth grade. The noncognitive measures are also allowed to depend on the cognitive measures. The skills have been normalized to have a variance of 1 separately for each gender. The whiskers represent the standard errors for each estimate, and the symbols on the bars indicate the results from tests of significance. The standard errors were calculated by using 400 bootstrap samples and allowed for clustering at the school-cohort level. There are 12,560 observations for males and 14,445 observations for females. Sources: ACS and OneGoal and CPS administrative data.

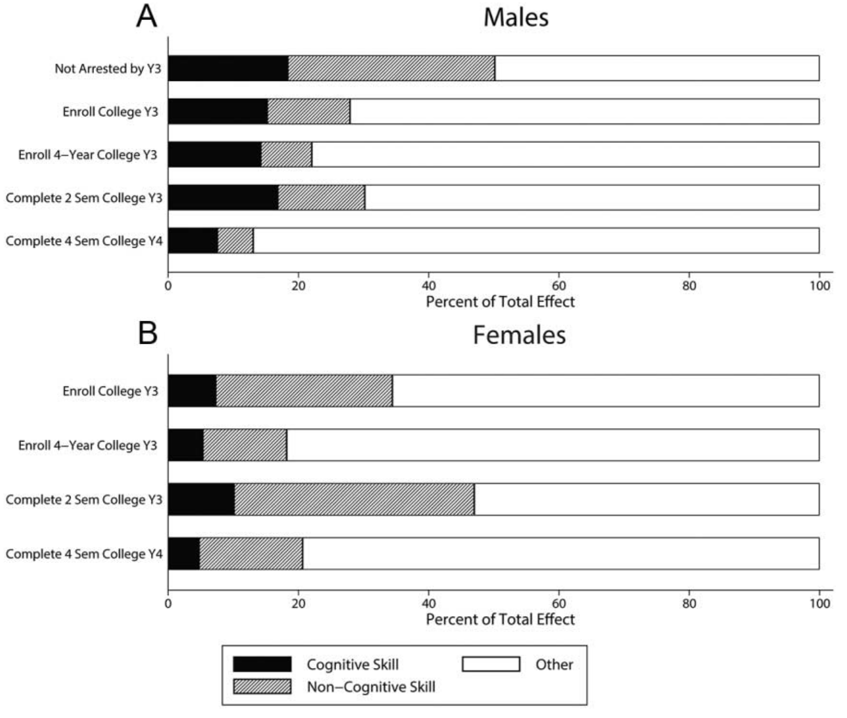

Figure 7.—Percentage of total effect due to cognitive skill, noncognitive skill, or other factors. Shown are the percentages of the total effect that can be attributed to improvements in cognitive skill, improvements in noncognitive skill, and other factors. Only outcomes with effects that are statistically different from zero are displayed. For males, the number of observations ranges from 6,930 to 9,301, depending on the availability of the outcome data. For females, the number of observations ranges from 8,481 to 10,905, depending on the availability of the outcome data. Sources: ACS and OneGoal, CPS, CPD, and NSC administrative data.

improvements in both cognitive and noncognitive skills account for part of the treatment effects. For arrests, these skills account for similar amounts of the treatment effect. For college indicators, changes in cognitive skills account for more of the treatment effect. For females, changes in cognitive skills explain almost none of the treatment effects. This result is consistent with figure 6, which shows that OneGoal had little effect on cognitive skills for females.

For both males and females, the "other factors" account for much of the treatment effect. These other factors might come from the information that OneGoal provides students about college enrollment, other forms of support, or changes in other types of skills. For example, students may have benefited from support with writing college essays or visiting college campuses. In addition, some of the skills that OneGoal aims to foster—such as teamwork skills—may not be captured by the measures used in this paper but may still relate to college outcomes. Nevertheless, these estimates suggest that providing mentorship and skill development also plays an important role.

VIII. Conclusion

We evaluate OneGoal, a program that attempts to help disadvantaged high school students complete college by improving noncognitive skills that are not captured by test scores. It teaches noncognitive skills through specific lessons and gives students a chance to apply those lessons to both schoolwork and the college application process.

We estimate that OneGoal reduces arrests by 5 percentage points for males and increases college enrollment by 10–20 percentage points for both males and females. Our panel is too short to estimate its effect on college graduation. Improvements in cognitive and noncognitive skill account for up to one-third of these effects. We find that OneGoal also improves outcomes through another factor, possibly the information that participants receive about applying to college. These results suggest that programs combining targeted information with skill development are promising. In addition, OneGoal may be a cost-effective way to improve these outcomes. During the time frame covered by this study, we estimate that OneGoal cost approximately $3,915 per student.[38]

These findings build on a growing body of evidence that suggests adolescent interventions can be effective. Some previous evidence suggested that early-childhood programs have been more cost-effective than adolescent programs.[39] This conclusion is partly an artifact of the types of adolescent interventions that have been studied.[40] For several adolescent programs, early evaluations suggested that the programs were successful, but longer follow-ups revealed that the effects faded, likely because they provided incentives that were tied to only short-term successes or temporarily modified the participants' environment.[41] Unlike these programs, OneGoal does not provide short-term incentives or drastically modify the students' environment.

OneGoal shares similarities with three promising types of adolescent interventions. The first type combines mentoring, work-based training, and a

[38] These costs include OneGoal's general operating costs of $1,492 per student, a stipend paid to teachers of $3,000 over the 3-year period, ACT preparation costs of $99 per student, and the cost of teachers' time to teach the class. The per student cost calculation is based on an assumption that OneGoal classes have an average of 25 students, teachers spend one-quarter of their class time teaching OneGoal students, and OneGoal teachers receive an annual salary of $73,486, consistent with CPS's proposed budget in 2012 (Chicago Public Schools 2012a). The cost estimates do not include deadweight costs of taxes.

[39] For evidence on successful early-childhood programs, see, e.g., Heckman et al. (2010), Reynolds et al. (2011), and Gertler et al. (2014).

[40] See Heckman and Kautz (2014) and Kautz et al. (2014) for reviews.

[41] Job Corps appeared to have short-term "incapacitation" effects on crime because it housed participants in a residential facility (Schochet, Burghardt, and McConnell 2008). The National Guard ChalleNGe program, another residentially based intervention for adolescents, also seemed to have similar incapacitation effects (Bloom, Gardenhire-Crooks, and Mandsager 2009; Millenky, Bloom, and Dillon 2010; Millenky et al. 2011). The Quantum Opportunity Program had a short-term effect on college enrollment, but it also provided large financial incentives (around $1,000) for participants to enroll in college (Rodríguez-Planas 2012).

curriculum that teaches specific noncognitive skills so that participants can immediately apply the skills they learn (Kemple and Snipes 2000; Kemple and Willner 2008; Roder and Elliot 2011, 2014; Fein et al. 2021; Katz et al. 2022). The second type provides adolescents and young adults with support in overcoming obstacles, with specific types of information, or with assistance at a time when it is particularly useful to them (e.g., information on how to complete financial aid forms or college applications; Bettinger et al. 2012; Carrell and Sacerdote 2013; Barr and Castleman 2021). The third type provides advising or case-management services once students are in college (Sommo et al. 2018; Weiss et al. 2019; Evans et al. 2020; Barr and Castleman 2021; Hallberg et al. 2023). Like these promising programs, OneGoal teaches noncognitive skills in the school setting, where students can apply the lessons immediately, provides specific information and assistance that is directly relevant to the process of selecting and applying to colleges, and offers ongoing support once students are in college.

In conducting the evaluation, we also showcase a way to measure noncognitive skills by use of administrative data available in most schools. This measure outperforms test scores in predicting arrests and high school graduation. Our method of measuring cognitive and noncognitive skills also minimized measurement error by using multiple measures that captured a common set of factors. Consequently, comparing outcomes of OneGoal participants and nonparticipants conditional on those measures of skills—instead of conditional on the variables that generated them as is typical when evaluating programs—is a more robust approach because measurement error in a conditioning variable can lead to bias in the estimated treatment effect.

Our evaluation demonstrates the importance of accounting for noncognitive skills. First, we show that before they enter the program, OneGoal participants tend to have higher levels of noncognitive skills than nonparticipants. If we did not account for these differences, we would overestimate the effects of OneGoal. Second, we find that OneGoal has a relatively small effect on test scores (cognitive skills) but that it has large effects on other outcomes, such as college enrollment. If we had measured only test scores and not noncognitive skills or other outcomes, we would have underestimated the effects of OneGoal. This evidence reveals the dangers of modern education policies that rely heavily on achievement test scores to assess students and schools.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge. 2022. "When Should You Adjust Standard Errors for Clustering?" *Q.J.E.* 138 (1): 1–35.

Ackerman, P. L., and E. D. Heggestad. 1997. "Intelligence, Personality, and Interests: Evidence for Overlapping Traits." *Psychological Bull.* 121 (2): 219–45.

ACT, Inc. 2007. *The ACT Technical Manual.* Iowa City, IA: ACT.

———. 2013a. *The ACT Explore Technical Manual.* Iowa City, IA: ACT.

———. 2013b. *The ACT Plan Technical Manual.* Iowa City, IA: ACT.

Almlund, M., A. Duckworth, J. J. Heckman, and T. Kautz. 2011. "Personality Psychology and Economics." In *Handbook of the Economics of Education*, vol. 4, edited by E. A. Hanushek, S. Machin, and L. Wößmann, 1–181. Amsterdam: North-Holland.

American Psychological Association. 2007. *APA Dictionary of Psychology*. Washington, DC: American Psychological Assoc.

Anderson, T. W., and H. Rubin. 1956. "Statistical Inference in Factor Analysis." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 5, edited by J. Neyman, 111–50. Berkeley: Univ. California Press.

Barr, A. C., and B. L. Castleman. 2021. "The Bottom Line on College Advising: Large Increases in Degree Attainment." EdWorkingPaper no. 21-481, Annenberg Inst. for School Reform, Brown Univ.

Benda, B. B. 2005. "The Robustness of Self-Control in Relation to Form of Delinquency." *Youth and Society* 36 (4): 418–44.

Bettinger, E. P., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *Q.J.E.* 127 (3): 1205–42.

Bloom, D., A. Gardenhire-Crooks, and C. L. Mandsager. 2009. "Reengaging High School Dropouts: Early Results of the National Guard Youth ChalleNGe Program Evaluation." Report, MDRC, New York.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel. 2008. "The Economics and Psychology of Personality Traits." *J. Human Resources* 43 (4): 972–1059.

Borghans, L., B. H. H. Golsteyn, J. Heckman, and J. E. Humphries. 2011. "Identification Problems in Personality Psychology." *Personality and Individual Differences* 51 (3): 315–20.

———. 2016. "What Grades and Achievement Tests Measure." *Proc. Nat. Acad. Sci. USA* 113 (47): 13354–59.

Carrell, S. E., and B. Sacerdote. 2013. "Late Interventions Matter Too: The Case of College Coaching in New Hampshire." Working Paper no. 19031 (May), NBER, Cambridge, MA.

Chicago Public Schools. 2002. "Uniform Discipline Code for Chicago Public Schools for 2002–2003 School Year." Board Report no. 02-0626-PO01, Chicago Public Schools, Chicago.

———. 2003. "Uniform Discipline Code for Chicago Public Schools for 2003–2004 School Year." Board Report no. 03-0723-PO02, Chicago Public Schools, Chicago.

———. 2004. "Uniform Discipline Code for Chicago Public Schools for 2004–2005 School Year." Board Report no. 04-0728-PO2, Chicago Public Schools, Chicago.

———. 2005. "Uniform Discipline Code: The Student Code of Conduct for All Chicago Public Schools Students for the 2005–2006 School Year." Board Report no. 05-0727-PO1, Chicago Public Schools, Chicago.

———. 2006a. "Student Code of Conduct (Formerly the Uniform Discipline Code) for Chicago Public Schools Effective September 12, 2006." Board Report no. 06-0823-PO3, Chicago Public Schools, Chicago.

———. 2006b. "Student Code of Conduct (Formerly the Uniform Discipline Code) for Chicago Public Schools for the 2006–2007 School Year." Board Report no. 06-0726-PO8, Chicago Public Schools, Chicago.

———. 2007. "Student Code of Conduct for Chicago Public Schools for the 2007–2008 School Year." Board Report no. 07-0627-PO5, Chicago Public Schools, Chicago.

———. 2008. "Student Code of Conduct for Chicago Public Schools for the 2008–2009 School Year." Board Report no. 08-0723-PO5, Chicago Public Schools, Chicago.

———. 2009. "Student Code of Conduct for Chicago Public Schools for the 2009–2010 School Year." Board Report no. 09-0722-PO5, Chicago Public Schools, Chicago.

———. 2010. "Student Code of Conduct for Chicago Public Schools for the 2010–2011 School Year." Board Report no. 10-0728-PO2, Chicago Public Schools, Chicago.

———. 2011. "Student Code of Conduct for Chicago Public Schools for the 2011–2012 School Year." Board Report no. 11-0727-PO2, Chicago Public Schools, Chicago.

———. 2012a. "Proposed Budget 2011–2012." Report, Chicago Public Schools, Chicago.

———. 2012b. "Student Code of Conduct for Chicago Public Schools for the 2012–2013 School Year." Board Report no. 12-0627-PO1, Chicago Public Schools, Chicago.

———. 2013. "Grade Point Average Policy." Board Report no. 13-0828-PO6, Chicago Public Schools, Chicago.

Clark, M. A., H. S. Chiang, T. Silva, S. McConnell, K. Sonnenfeld, A. Erbe, M. Puma, E. Warner, and S. Schmidt. 2013. "The Effectiveness of Secondary Math Teachers from Teach for America and the Teaching Fellows Programs." NCEE 2013–4015, Nat. Center for Educ. Evaluation and Regional Assistance, Inst. Educ. Sci., US Dept. Educ., Washington, DC.

Cook, P. J., K. Dodge, G. Farkas, R. G. Fryer Jr., J, Guryan, J. Ludwig, S. Mayer, H. Pollock, and L. Steinberg. 2014. "The (Surprising) Efficacy of Academic and Behavioral Intervention with Disadvantaged Youth: Results from a Randomized Experiment in Chicago." Working Paper no. 19862 (January), NBER, Cambridge, MA.

Cunha, F., and J. J. Heckman. 2007. "The Technology of Skill Formation." *A.E.R.* 97 (2): 31–47.

———. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *J. Human Resources* 43 (4): 738–82.

Cunha, F., J. J. Heckman, and S. M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.

de Chaisemartin, C., and X. d'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *A.E.R.* 110 (9): 2964–96.

Decker, P. T., D. P. Mayer, and S. Glazerman. 2004. "The Effects of Teach for America on Students: Findings from a National Evaluation." Report, Mathematica, Princeton, NJ.

Duckworth, A. L., P. D. Quinn, and E. Tsukayama. 2012. "What No Child Left Behind Leaves Behind: The Roles of IQ and Self-Control in Predicting Standardized Achievement Test Scores and Report Card Grades." *J. Educ. Psychology* 104 (2): 439–51.

Evans, W. N., M. S. Kearney, B. Perry, and J. X. Sullivan. 2020. "Increasing Community College Completion Rates among Low-Income Students: Evidence from a Randomized Controlled Trial Evaluation of a Case-Management Intervention." *J. Policy Analysis and Management* 39 (4): 930–65.

Fein, D., S. Dastrup, and K. Burnett. 2021. "Still Bridging the Opportunity Divide for Low-Income Youth: Year Up's Longer-Term Impacts." OPRE Report no. 2021–56, Office of Planning, Res., and Evaluation, Washington, DC.

Gertler, P., J. J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344 (6187): 998–1001.

Hallberg, K., C. Eaglin, R. Hirami, and A. Ruizesparze. 2022. "Supporting Students in Achieving Their Postsecondary Goals: A Quasi-Experimental Program Evaluation of OneGoal." Report, Inclusive Economy Lab, Univ. Chicago.

Hallberg, K., K. Hofmeister, M. Bertrand, and B. Morgan. 2023. "Supporting Community College Student Success: Evidence from a Randomized Controlled Trial." *J. Res. Educ. Effectiveness* 16 (1): 63–81.

Heckman, J. J., J. E. Humphries, and T. Kautz, editors. 2014. *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago: Univ. Chicago Press.

Heckman, J. J., J. E. Humphries, and G. Veramendi. 2014. "Education, Health and Wages." Working Paper no. 19971 (March), NBER, Cambridge, MA.

———. 2018. "Web Appendix for the Non-market Benefits of Education and Ability." Manuscript, Univ. Chicago.

Heckman, J. J., T. Jagelka, and T. Kautz. 2021. "Some Contributions of Economics to the Study of Personality." In *Handbook of Personality*, edited by O. P. John and R. W. Robins, 853–92. New York: Guilford.

Heckman, J. J., and T. Kautz. 2012. "Hard Evidence on Soft Skills." *Labour Econ.* 19 (4): 451–64.

———. 2014. "Fostering and Measuring Skills: Interventions That Improve Character and Cognition." In *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, edited by J. J. Heckman, J. E. Humphries, and T. Kautz, 341–430. Chicago: Univ. Chicago Press.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz. 2010. "The Rate of Return to the HighScope Perry Preschool Program." *J. Public Econ.* 94 (1–2): 114–28.

Heckman, J. J., and S. Navarro-Lozano. 2004. "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models." *Rev. Econ. and Statis.* 86 (1): 30–57.

Heckman, J. J., and R. Pinto. 2015. "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs." *Econometric Rev.* 34 (1–2): 6–31.

Heckman, J. J., and E. J. Vytlacil. 2007. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments." In *Handbook of Econometrics*, vol. 6B, edited by J. Heckman and E. Leamer, chap. 7. Amsterdam: North-Holland.

Heller, S. B., A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, and H. A. Pollack. 2017. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." *Q.J.E.* 132 (1): 1–54.

Horn, J. L. 1965. "A Rationale and Test for the Number of Factors in Factor Analysis." *Psychometrika* 30 (2): 179–85.

Jackson, C. K. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes." *J.P.E.* 126 (5): 2072–107.

Kaiser, H. F. 1960. "The Application of Electronic Computers to Factor Analysis." *Educ. and Psychological Measurement* 20 (1): 141–51.

Katz, L. F., J. Roth, R. Hendra, and K. Schaberg. 2022. "Why Do Sectoral Employment Programs Work? Lessons from WorkAdvance." *J. Labor Econ.* 40 (S1): S249–S291.

Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans. 2014. "Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success." Report, Org. Econ. Co-operation and Development, Paris.

Kemple, J., and J. C. Snipes. 2000. "Career Academies: Impacts on Student Engagement and Performance in High School." Report, MDRC, New York. http://www.mdrc.org/sites/default/files/full_45.pdf.

Kemple, J., and C. Willner. 2008. "Career Academies: Long-Term Impacts on Labor Market Outcomes, Educational Attainment, and Transitions to Adulthood." Report, MDRC, New York. http://www.mdrc.org/sites/default/files/full_50.pdf.

Lira, B., J. M. O'Brien, P. A. Peña, et al. 2022. "Large Studies Reveal How Reference Bias Limits Policy Applications of Self-Report Measures." *Scientific Reports* 12 (1): 1–12.

Maloney, M., S. Mahaffie, G. Hernandez, A. Usher, and J. Nagaoka. 2021. "The Educational Attainment of Chicago Public Schools Students: 2020." Report, Consortium on Chicago School Res., Chicago.

Millenky, M., D. Bloom, and C. Dillon. 2010. "Making the Transition: Interim Results of the National Guard Youth ChalleNGe Evaluation." Report, MDRC, New York. http://www.mdrc.org/sites/default/files/full_434.pdf.

Millenky, M., D. Bloom, S. Muller-Ravett, and J. Broadus. 2011. "Staying on Course: Three-Year Results of the National Guard Youth ChalleNGe Evaluation." Report, MDRC, New York. http://www.mdrc.org/sites/default/files/full_510.pdf.

Pratt, T. C., and F. T. Cullen. 2000. "The Empirical Status of Gottfredson and Hirschi's General Theory of Crime: A Meta-analysis." *Criminology* 38 (3): 931–64.

Reynolds, A. J., J. A. Temple, S.-R. Ou, I. A. Arteaga, and B. A. B. White. 2011. "School-Based Early Childhood Education and Age-28 Well-Being: Effects by Timing, Dosage, and Subgroups." *Science* 333 (6040): 360–64.

Roberts, B. W., N. R. Kuncel, R. L. Shiner, A. Caspi, and L. R. Goldberg. 2007. "The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes." *Perspectives Psychological Sci.* 2 (4): 313–45.

Roder, A., and M. Elliot. 2011. "A Promising Start: Year Up's Initial Impacts on Low-Income Young Adults' Careers." Report, Econ. Mobility Corporation, New York.

———. 2014. "Sustained Gains: Year Up's Continued Impact on Young Adults' Earnings." Report, Econ. Mobility Corporation, New York.

Rodríguez-Planas, N. 2012. "Longer-Term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the United States." *American Econ. J. Appl. Econ.* 4 (4): 121–39.

Schmitt, D. P., J. Allik, R. R. McCrae, and V. Benet-Martínez. 2007. "The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description across 56 Nations." *J. Cross-Cultural Psychology* 38 (2): 173–212.

Schochet, P. Z., J. Burghardt, and S. McConnell. 2008. "Does Job Corps Work? Impact Findings from the National Job Corps Study." *A.E.R.* 98 (5): 1864–86.

Skrondal, A., and P. Laake. 2001. "Regression among Factor Scores." *Psychometrika* 66 (4): 563–76.

Sommo, C., D. Cullinan, M. Manno, S. Blake, and E. Alonzo. 2018. "Doubling Graduation Rates in a New State: Two-Year Findings from the ASAP Ohio Demonstration." Report, MDRC, New York.

Tough, P. 2012. *How Children Succeed: Grit, Curiosity, and the Hidden Power of Character.* Boston: Houghton Mifflin Harcourt.

Tyler, R. W. 1940. "The Place of Evaluation in Modern Education." *Elementary School J.* 41 (1): 19–27.

Weiss, M. J., A. Ratledge, C. Sommo, and H. Gupta. 2019. "Supporting Community College Students from Start to Degree Completion: Long-Term Evidence from a Randomized Trial of CUNY's ASAP." *American Econ. J. Appl. Econ.* 11 (3): 253–97.

West, M. R., M. A. Kraft, A. S. Finn, R. E. Martin, A. L. Duckworth, C. F. O. Gabrieli, and J. D. E. Gabrieli. 2016. "Promise and Paradox: Measuring Students' Noncognitive Skills and the Impact of Schooling." *Educ. Evaluation and Policy Analysis* 38 (1): 148–70.

Williams, B. 2020. "Identification of the Linear Factor Model." *Econometric Rev.* 39 (1): 92–109.

Yeager, D. S. 2017. "Social and Emotional Learning Programs for Adolescents." *Future of Children* 27 (1): 73–94.