Measuring and Fostering Non-Cognitive Skills in Adolescence: Evidence from Chicago Public Schools and the OneGoal Program

Tim Kautz and Wladimir Zanoni *

This draft, November 17, 2014.

*Kautz: Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago IL 60637; email: tkautz@uchicago.edu. Zanoni: Chapin Hall at the University of Chicago, 1313 E. 60th Street, Chicago, IL 60637; email: wzanoni@chapinhall.org. 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 report are those of the authors and do not necessarily reflect the views of any of the funders. We thank the OneGoal staff and especially Kylie Vadnais for continual assistance in understanding the OneGoal program and data. We thank the Absences and Truancy Department of the Chicago Public Schools for guidance in understanding school policies. We benefited from helpful comments from Elaine Allensworth, Enghin Atalay, Stephane Bonhomme, Kathleen Danna, Angela Duckworth, Philipp Eisenhauer, Jorge Luis García. Miriam Gensowski, Shannon Guiltinan, Eric Hansford, Masha Hansford, Dana Hepper, John Eric Humphries, Richard Kautz, Gilbert Lemieux, Magne Mogstad, María Rosales-Rueda, Todd Rosencranz, Krishanu Sengupta, Bradley Setzler, Lauren Sompayrac, Melissa Tartari, Jake Torcasso, Danielle Tyree, Sarah Vernon, and the participants of the Lifecycle Working Group and Family Economics Workshop at the University of Chicago. We thank Peter Porcino 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.

Abstract

Recent evidence has established that *non-cognitive* skills (e.g., persistence and self-control) are valuable in the labor market and are malleable throughout adolescence. Some recent high school interventions have been developed to foster these skills, but there is little evidence on whether they are effective. Using administrative data, we apply two methods to evaluate an intervention called OneGoal, which attempts to help disadvantaged students attend and complete college in part by teaching non-cognitive skills. First, we compare the outcomes of participants and non-participants with similar pre-program cognitive and non-cognitive skills. In doing so, we develop and validate a measure of non-cognitive skill that is based on readily available data and rivals standard measures of cognitive skill in predicting educational attainment. Second, we use an instrumental variable difference-in-difference approach that exploits the fact that OneGoal was introduced into different schools at different times. We estimate that OneGoal increases college enrollment by 10–20 percentage points. We demonstrate that improvements in non-cognitive skill account for 15–30 percent of the effect.

Contents

1 Introduction

Many disadvantaged high school students have poor life outcomes. For example, data from the Chicago Public Schools (CPS) show that about 60% of entering 9th graders graduate high school within five years and only about 15% will earn college degree. Most educational improvement strategies focus on cognitive skills as measured by achievement test scores. However, test scores miss *non-cognitive* 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 at later ages (Heckman and Kautz, 2014; Kautz et al., 2014). This paper shows that a high school intervention can improve outcomes by developing non-cognitive skills and that non-cognitive skills can be measured using administrative data that is readily available from school records.

The main challenge in evaluating OneGoal is to account for the possibility of selection bias. Participants might have fared better than non-participants even in the absence of the program. We address the selection problem in two ways. First, we adopt a matching approach in which we compare the outcomes of OneGoal participants with those of other CPS students who have similar pre-program levels of cognitive and non-cognitive skills. We show that accounting for non-cognitive skill is important because more motivated students are more likely to participate in OneGoal.

Our second approach exploits the fact that OneGoal was introduced into different schools at different times. As a result, there are cohorts of students who were ineligible to participate in OneGoal simply because it was not offered in their school. We use eligibility as an instrumental variable to compare students who were eligible to participate with those that were not, similar to comparing the treatment and control groups in a randomized experiment. We use a difference-in-difference specification that accounts for differences in baseline outcomes between schools. Under both of these approaches, meaningful effects of OneGoal are identified and can be estimated non-parametrically.

Our first approach requires us to proxy unobserved cognitive and non-cognitive skills. Our measure of cognitive skill is based on students' achievement test scores, while our measure of non-cognitive skill is based on their grades, credits, disciplinary infractions, and absences. Our procedure accounts for measurement error and removes the cognitive component from the non-cognitive measures, so that non-cognitive skills are defined relative to test scores. In this framework, a student with low test scores but average grades would still have a high level of non-cognitive skill. We validate this relatively novel[1] measure of non-cognitive skills by demonstrating that it predicts outcomes that matter. For a sample of ninth graders, this measure of non-cognitive skill explains 30% of the variation in high school graduation, while standard measures of cognitive skill explain 10%.

We implement these methods using large, detailed, and new dataset that we construct by combining data from the Chicago Public Schools, the Chicago Police Department (CPD), OneGoal, the National Student Clearinghouse (NSC), and the American Community Survey (ACS). This novel dataset has over 50,000 observations and contains information throughout high school and college.

OneGoal appears to be a successful intervention. After accounting for selection effects, we find that OneGoal improves intermediate outcomes in high school such as grades, days absent, test scores, and credits earned. More importantly, we estimate that OneGoal reduces arrests for males by 5 percentage points and increases both college enrollment and college persistence by 10–20 percentage points for both males and females. We find similar results from both of our approaches, but estimates from the instrumental variable approach are less precise.

We subject these estimates to a battery of sensitivity tests. We find very similar results regardless of the exact specification or method that we employ. Using unique features of the data, we conduct a test of the assumptions justifying matching that further validates our

[1] See Heckman et al. (2012) and Jackson (2013) for recent papers in economics that use similar measures. See Duckworth et al. (2012) and Borghans et al. (2011) for evidence that grades are related to standard measures of non-cognitive skills.

measures of skill. We also offer several types of evidence that suggest that the difference-in-difference method is valid; that is, eligibility is related to outcomes only insofar as it affects participation in OneGoal.

In three ways, this paper contributes to knowledge about adolescent interventions. First, when compared to evaluations of other adolescent programs, it has a relatively long follow-up and considers a broader set of outcomes.[2] Second, of the adolescent interventions that are well studied, OneGoal is one of the few that is successful.[3] Third, we dig deeper than most evaluations of adolescent programs by demonstrating that the improvements in outcomes are linked to improvements in skills.

Our analysis also illustrates broader points about test scores and non-cognitive skills. Before entering the program, OneGoal participants tend to have near-average cognitive skills (test scores) but above-average non-cognitive skills compared to their peers. If we had not accounted for differences in non-cognitive skills, we would have overestimated the effects of OneGoal on college outcomes by 30%–40%. 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 the program has the largest effects on outcomes other than test scores, such as college enrollment. We further demonstrate that OneGoal improves these outcomes in part because it improves students' non-cognitive skills. These results reveal the dangers of evaluations (or other studies) that rely solely on test scores to measure skills.

Our analysis is relevant to evaluations of public policies. Some policy makers would like to move beyond using only test scores to assess students, teachers, and schools, and are searching for viable assessments of non-cognitive skills.[4] We validate one measure of non-cognitive skill based on data that are commonly collected by schools but rarely used to measure skills. This measure rivals and often outperforms achievement test scores in predicting arrests, high

[2]See the discussion in Heckman and Kautz (2014) and Table W3 in Web Appendix Section W2.3, which summarizes the nature and efficacy of a wide range of interventions.

[3]See also Heckman and Kautz (2014) and Table W3 in Web Appendix Section W2.3.

[4]For example, a group of California school districts (representing over one million students) has applied for a No Child Left Behind waiver in part to incorporate "yet to be determined" measures of "non-cognitive" skills into their school assessments (CORE, 2013).

school graduation, college enrollment, and college graduation.

The paper proceeds as follows. Section 2 provides a description of the OneGoal program and reviews the literature on skill measurement and adolescent interventions. Section 3 defines the treatment effects and discusses identification. Section 4 describes our data and validates our measures of skill. Section 5 provides our main analysis, which includes a comparison of OneGoal participants to their peers, estimates of the treatment effects, a mediation analysis, and an investigation of treatment effect heterogeneity. Section 6 presents our sensitivity analyses. Section 7 concludes the paper.

2 Background

2.1 OneGoal

In this section, we summarize the OneGoal approach. Because accounting for selection is the main challenge of this paper, 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. 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 course work and other challenges.

The other half of the curriculum provides lessons on specific non-cognitive skills and gives OneGoal Fellows an opportunity to apply the lessons to their school work and the college admissions process. For example, one lesson covers how to set goals and create an action plan to accomplish those 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 a teacher (the Program Director) by checking references, interviewing teachers, and observing them in the classroom.[5] 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)[6] and their commitment to the application process. For these reasons, OneGoal might select more motivated students with higher non-cognitive skills. In Section 5, we show that compared to their peers in CPS, OneGoal Fellows have near-average cognitive skills but above-average non-cognitive skills, but even after accounting for these differences, we find that OneGoal is effective.

2.2 Review of the Literature on Interventions

In this section, we compare OneGoal with other adolescent interventions. The comparisons suggest why OneGoal might be effective while other interventions fail.

A growing body of evidence suggests that early childhood programs have been more cost-effective than adolescent programs.[7] This conclusion is partly an artifact of the types of adolescent interventions that have been studied.[8] Two of the best-studied adolescent interventions are Job Corps and the Quantum Opportunity Program (QOP). For both programs, early evaluations suggested that the programs were successful, but longer follow-

[5]See Table W1 in Web Appendix Section W2.1 for more details on the teacher recruitment process.

[6]See Figure W1 in Section W2.1 of the Web Appendix for how students were assessed on the five leadership principles.

[7]For evidence on successful early childhood programs see, for example, Heckman et al. (2010), Gertler et al. (2013), and Reynolds et al. (2011).

[8]See Heckman and Kautz (2014) and Kautz et al. (2014) for reviews. See also Table W3 in Section 2.2 of the Web Appendix.

ups revealed that the effects faded, likely because both programs provided incentives that were tied to only short-term successes.[9] A number of other programs show promise but their evaluations are too short for strong conclusions to be drawn.[10]

Recent evaluations suggest two promising types of adolescent interventions. The first type combines mentoring, work-based training, and a curriculum that teaches specific non-cognitive skills. These programs have strong effects on labor market outcomes.[11] They emphasize skills such as punctuality, teamwork, and discipline and give participants a chance to apply these skills in a work environment.

The second type provides adolescents and young adults with specific types of information or assistance at a time when it is particularly useful to them. For example, the Dartmouth College Coaching Program is a relatively short-term intervention that provides students with information about college applications and helps them complete applications (Carrell and Sacerdote, 2013). The program increased college enrollment of female students by about 15 percentage points but had no effect on male students. Similarly, Bettinger et al. (2012) find that providing families with information on how to complete financial aid forms can increase college enrollment by up to 8 percentage points.

OneGoal's curriculum incorporates components of both of these models. It teaches non-cognitive skills in the school setting, where students can apply the lessons immediately, and it provides specific information and assistance that is directly relevant to the process of selecting and applying to colleges. Unlike Job Corps and QOP, OneGoal does not provide short-term incentives or drastically modify the students' environment.

[9]Job Corps appeared to have short-term "incapacitation" effects on crime because it housed participants in a residential facility (Schochet et al., 2008). The National Guard ChalleNGe program, another residentially based intervention for adolescents, also seemed to have similar incapacitation effects (Bloom et al., 2009; Millenky et al., 2010, 2011). QOP 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).

[10]See, for example, Big Brothers Big Sisters (Tierney et al., 1995; Aos et al., 2004), Becoming a Man (BAM) (Cook et al., 2014), and the National Guard ChalleNGe Program (Bloom et al., 2009; Millenky et al., 2010, 2011).

[11]See Career Academies (Kemple and Snipes, 2000; Kemple and Willner, 2008) and the Year-Up program (Roder and Elliot, 2011, 2014). While the longest evaluation of Year-Up is only three years, it seems especially promising because it has increased the hourly wages of participants, suggesting that they have gained skills that are valued in the labor market.

2.3 Review the Literature on Skill Measurement

In this section, we discuss the advantages of using administrative data to measure non-cognitive skills. Psychologists typically elicit personality traits (non-cognitive 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?" Recently, economists have argued that it is valid to measure non-cognitive skills using a broader 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 non-cognitive skills using grades, absences, disciplinary infractions, and credits earned. These measures are valid because they depend on skills beyond raw smarts. For example, earning course credits requires showing up to class and completing assignments.[12]

Some psychologists have argued that it is tautological to use this approach because it uses behavior to predict future behaviors.[13] 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, as they require respondents to fill out questionnaires (which is itself a behavior) or report the types of behaviors that they tend to exhibit. The real question is whether the measure predicts outcomes that matter. Our measure of non-cognitive skills is more predictive of life outcomes than what is typically found for self-reported measures.[14]

Chicago Public Schools has also implicitly started to adopt this broader approach by supplementing test scores with the "on-track" indicator based on school credits attained (Al-

[12]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).This logic has been fruitfully applied by Heckman et al. (2012) and Jackson (2013), who measure non-cognitive skills using adolescent risky behaviors and data from school transcripts.

[13]See, for example, the discussion in Pratt and Cullen (2000) and Benda (2005).

[14]See Almlund et al. (2011) for a review of studies that use self-reported measures. Our measures might be more predictive because they avoid a problem known as "reference bias," which arises in self-reported questionnaires when respondents rate themselves in comparison to their peers rather than to the whole population.For a discussion of reference bias and further examples, see Heckman and Kautz (2014), Schmitt et al. (2007), and Duckworth (2012).

lensworth and Easton, 2005). Allensworth and Easton (2007) show that the on-track indicator is predictive of eventual high school graduation. We build on this measure by using an even broader array of measures to account for measurement error.

3 Defining and Identifying Treatment Effects

In this section, we define treatment effects and discuss identification. We adopt a general framework to highlight the tradeoffs between different models.

3.1 Policy Questions and Treatment Effects

We first define treatment effects that directly correspond to policy questions. We adopt a standard potential outcomes framework. For each student, define $D = 1$ if they would, given the opportunity, choose to participate in OneGoal and $D = 0$ if they would choose not participate in OneGoal. Let Y_1 be an outcome if they were to participate and Y_0 be their outcome if not. Let X be a vector of observed covariates (i.e., basic demographics), and let θ be a set of unobserved skills (e.g., cognitive and non-cognitive skills) that could be proxied using data.

We consider two primary treatment effects: the average treatment effect (ATE) and the effect of treatment on the treated (TT). We allow for the possibility that these effects depend on the observed covariates and skills:

$$\begin{aligned} ATE\left(X,\theta\right) &= E\left[Y_1 - Y_0 | X, \theta\right], \\ TT\left(X,\theta\right) &= E\left[Y_1 - Y_0 | D = 1, X, \theta\right]. \end{aligned}$$

These treatment effects address specific policy questions that are relevant to OneGoal and other adolescent programs. ATE is the effect of OneGoal if it were to be applied to the whole population (all students in CPS). This parameter is relevant because aspects of the OneGoal curriculum could conceivably be made standard for all eleventh and twelfth graders. For example, Uplift Community High School has already adopted a version of OneGoal for all of its students. The TT is the effect for students who choose to enroll under the current model and corresponds to the treatment effect as OneGoal is implemented now. We allow these effects to depend on pre-program characteristics in order to identify the types of students who would benefit most from OneGoal or similar programs. In Section 5.4, we provide some evidence that students with low pre-program cognitive skills benefit most.

3.2 Identification

In this section, we discuss the main assumptions underlying our two identification strategies. Similar to Heckman and Navarro-Lozano (2004) and Heckman and Vytlacil (2007a), we discuss identification in the context of a model of economic choice. Assume that Y_k can be decomposed into a function $\mu_k(X, \theta)$ and a separable error term U_k that also depends on treatment:

$$Y_1 = \mu_1(X, \theta) + U_1,$$
$$Y_0 = \mu_0(X, \theta) + U_0.$$

We define a variable specific to our application, Z, which affects the possibility of receiving treatment. In our application, $Z = 1$ when a student is eligible for OneGoal in the sense that they are in a OneGoal school during a period when OneGoal is offered for their cohort, and $Z = 0$ if they are not eligible. Z plays a crucial role in our second approach. Define the net benefit of participating in OneGoal as D^*, which also depends on X and θ. Let a A be an indicator for whether a student actually participates in OneGoal; that is, OneGoal is offered in their school ($Z = 1$) and they would choose to participate ($D = 1$). Students select treatment if the benefits outweigh the costs:

$$D^* = \mu_D(X,\theta) + U_D,$$

$$D = \begin{cases} 1 & \text{if } D^* \geq 0, \\ 0 & \text{if } D^* < 0, \end{cases}$$

$$A = \begin{cases} 1 & \text{if } D = 1 \text{ and } Z = 1, \\ 0 & \text{if } D = 0 \text{ or } Z = 0. \end{cases}$$

We adopt two approaches.

Approach 1: Matching on Demographics and Skills

Our main approach is to proxy the unobserved skills (θ) and compare the outcomes of OneGoal participants to those of other CPS students. This approach relaxes the typical matching assumption by allowing $(U_1, U_0) \not\perp\!\!\!\perp U_D | X$, where $\perp\!\!\!\perp$ denotes statistical independence. The main assumptions under this approach are

(M-1) $(U_1, U_0) \perp\!\!\!\perp U_D | X, \theta$,

(M-2) $0 < \Pr(D = 1 | X, \theta) < 1$.

Under conditions (M-1) and (M-2), ATE, $ATE(X,\theta)$, TT, and $TT(X,\theta)$ are non-parametrically identified over the support of (X,θ) and $ATE(X,\theta) = TT(X,\theta)$.[15] Under these assumptions the observed covariates (X) and skills (θ) account for the dependence between the decision to enter OneGoal and the outcomes.

Approach 2: Using OneGoal Eligibility as an Instrument

We supplement our main approach with an instrumental variable method that accounts for potential shortcomings of the matching approach. OneGoal was introduced to different

[15] See Heckman and Navarro-Lozano (2004) for a discussion.

schools at different times, so some students were ineligible to participate simply because OneGoal was not available to them at the time. We compare the outcomes of the eligible students to those of the ineligible students.

This comparison is analogous to a randomized experiment that gives the treatment group the option to take up treatment and ensures that the control group does not take up treatment. In these circumstances, randomization serves as a valid instrument for accepting the treatment and the instrumental variable estimator consistently estimates the treatment on the treated parameter. Our approach follows a similar logic and assumes

(E-1) $Z \perp\!\!\!\perp (U_1, U_0, U_D) \,|X, \theta$

(E-2) $\Pr(D = 1|X, \theta, Z = 1) = \Pr(D = 1|X, \theta, Z = 0)$.

Assumption (E-1) states that OneGoal eligibility is independent of outcomes and the decision to participate conditional on the covariates and skills. Assumption (E-2) is an "invariance" assumption that states that eligibility does not affect whether a student would like to participate. The standard instrumental variables estimator converges to the Wald estimand:

$$\frac{E[Y|X, \theta, Z = 1] - E[Y|X, \theta, Z = 0]}{E[A|X, \theta, Z = 1] - E[A|X, \theta, Z = 0]}.$$

Under assumptions (E-1) and (E-2), this expression simplifies to $E[Y_1 - Y_0|X, \theta, D = 1] = TT(X, \theta)$.[16] This approach accounts for the possibility that the matching assumptions fail; that is, X and θ do not capture all of the dependence between OneGoal participation and the outcomes so that $(U_1, U_0) \not\perp\!\!\!\perp U_D|X, \theta$. There are two possible cases: (1) there are additional unobserved skills that affect selection into OneGoal and also affect the outcomes (e.g., $U_1 = U_0 = U$ and $U \not\perp\!\!\!\perp U_D$) or (2) students select into OneGoal based on unobserved gains ($U_1 - U_0 \not\perp\!\!\!\perp U_D$). In either case, the treatment on the treated parameter is identified.

[16]See Heckman and Vytlacil (2007b) for a derivation and discussion in the case of experiments.To see this, note that based on the property of Z, $E[A|X, \theta, Z = 0] = 0$ and $E[Y|X, \theta, Z = 0] = E[Y_0|X, \theta, Z = 0]$. Under (E-1), $E[Y|X, \theta, Z = 1] = E[Y_1|D = 1]E[D] + E[Y_0|D = 0](1 - E[D])$. Plugging these expressions into the Wald estimand and rearranging terms yields the result.

Identification of the Distribution of θ

For the identification of the treatment effects, we have assumed that θ is observable. However, θ is not observable, and so we proxy θ. We apply the ideas discussed in Section 2.3 by using a broad set of behaviors to identify the distribution of cognitive and non-cognitive skills. This approach assumes that a large set of measures depend on a low-dimensional set of underlying latent variables. Using the covariance between measures, it is possible to identify the distribution of the latent variables.

In this paper, we assume that measures M_j depend on two latent variables that represent cognitive skills (θ_C) and non-cognitive 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 S_j.[17] In our application, for example, we allow attendance to depend on the distance a student lives from school. We assume a linear form:

$$M_j = \alpha_{C,j}\theta_C + \alpha_{N,j}\theta_N + \beta_j S_j + \varepsilon_j,$$

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

We set the scale of the latent variables by assuming that for one measure (k) the factor loading on cognitive skill is one ($\alpha_{C,k} = 1$) and that for another measure (l) the factor loading on non-cognitive skill is one ($\alpha_{N,l} = 1$).[18] We make additional assumptions so that the factors have a clear interpretation. In this paper, our measures are absences, grade point average, credits, disciplinary infractions, and subscores on achievement tests. We assume that the subscores on the achievement tests only depend on cognitive skill, so $\alpha_{N,j} = 0$ for them, whereas all the remaining measures depend on both cognitive and non-cognitive skills. We

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

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

additionally assume that the factors are statistically independent: $\theta_C \perp\!\!\!\perp \theta_N$.[19]

With these assumptions, the cognitive skill factor represents what is measured by achievement tests (after correcting for measurement error) and the non-cognitive skill factor represents what is captured by the other ninth-grade measures that is not explained by cognitive skill. Therefore, any predictive power of non-cognitive 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. This factor model is identified if there are at least two measures of cognitive skill and three measures that depend on both cognitive and non-cognitive skills.[20]

4 Data

4.1 Data

An important contribution of this paper is to merge data from five sources: OneGoal administrative records, Chicago Public Schools (CPS), the Chicago Police Department (CPD), the National Student Clearinghouse (NSC), and the American Community Survey (ACS). We merge these datasets together to construct histories of each CPS student who was in ninth or tenth grade between 2003 and 2013.[21] Of the 2,376 students accepted into OneGoal, we matched 2,342 (99%) of them with the CPS data.[22] Table 1 summarizes and defines the variables used in this study.

Since 2003, CPS has collected detailed administrative data on grade point averages

[19]The model is identified even if $\theta_C \not\perp\!\!\!\perp \theta_N$. In a sensitivity analysis we relax this assumption and cannot reject that $\theta_C \perp\!\!\!\perp \theta_N$. See Section 4.2 for a discussion.

[20]Anderson and Rubin (1956) and Williams (2012) show general conditions under which linear factor models are identified. This model satisfies those conditions. Web Appendix Section W3.3 presents an algebraic proof for this specialized case.

[21]See Web Appendix Section W5 for a more detailed description of the data and how we standardized the variables over time.

[22]See Table W2 in Web Appendix Section W2.1 for the number of students in each cohort in each school.

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
%tile Absences	Percentile ranking of total absences, standardized by grade and year	CPS
GPA	Grade point average, measured on a four-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 white, black, Hispanic, or other	CPS
High School Enrollment Status	Whether a student is actively enrolled, left as a non-graduate, 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
% of Single-Parent Households	Percent of single-parent households in a student's census block group	ACS, CPS
Employment Rate (Age 16–19)	Fraction of residents age 16–19 that are employed in a student's census block group	ACS, CPS
Enrollment Rate (Age 16–19)	Fraction of residents age 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

(GPAs);[23] absences; disciplinary infractions; ninth-, tenth-, and eleventh-grade test scores (Explore, Plan, and the ACT);[24] high school graduation status; student addresses; school addresses; special education status; race; gender; and age.[25]

Measuring 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.[26]

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.[27] 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). Due to the limited numbers 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 their school. We also use the addresses to identify each student's census block group (neighborhood), on which additional data are collected by the United States Census Bureau.[28] We link the student's census block group to neighborhood information from the ACS. These neighborhood characteristics supplement the control variables available in the CPS data.

We also use Chicago Policy Department (CPD) data that have been linked to students in

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

[24]For a description of these tests, see ACT, Inc. (2013a), ACT, Inc. (2013b), and ACT (2007).

[25]See Section W5.2 of the Web Appendix for a detailed discussion of how we standardized these variables over time. Some of these variables have been collected for longer periods of time.

[26]See Section W5.2 of the Web Appendix for a detailed description of how this change affected the measurement of absences and how we address these issues.

[27]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,b, 2007, 2008, 2009, 2010, 2011, 2012).

[28]See Section W5.3 of the Web Appendix for details on how the distances were calculated and block groups were assigned.

CPS. The arrest database contains all arrest records since 1999 that occurred in Chicago.

Data on post-secondary educational attainment comes from the National Student Clearinghouse (NSC).[29] For each student who completes high school or earns an alternative diploma, we access their 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.

4.2 Validating the Measurement System

Our matching approach relies on measuring the skills that both affect selection into OneGoal and future outcomes. Based on OneGoal's recruitment strategy, non-cognitive skills likely play a role in determining who participates (see Section 2.1). Because our measures of non-cognitive skill are vital to our study, we validate them by exploring the extent to which they predict outcomes. To do this, we consider whether ninth-grade measures predict outcomes in later years by analyzing a cohort of students who were first-time ninth graders in the fall of 2003. We first present simple correlations, and then we use a factor model that reduces the dimensionality of the data and corrects for measurement error.

Correlational Evidence

Figure 1 shows the correlation between ninth-grade test scores, GPAs, credits earned, absences, and disciplinary infractions. All variables are scaled so that a higher value corresponds to a more beneficial outcome. Two conclusions emerge. First, all of the variables are intercorrelated, suggesting that they measure something in common. Second, the extent to which they are correlated varies greatly, suggesting that they cannot be explained by a single underlying variable. For example, discipline and absences are highly correlated with each other, but both have a low correlation with achievement tests. GPA is relatively highly correlated with all measures.

[29]See Section W5.5 of the Web Appendix for a description of how we cleaned the NSC data.

Table 2 shows the predictive validity (R^2) from regressions of each measure on the outcomes in the first column. The last column shows the R^2 from using all measures. Test scores are a relatively poor predictor for many outcomes. For example, test scores predict only about 10% of the variation in completing high school, whereas absences and GPA predict about 20% and 30%. There is one exception: Test scores are by far the best predictors of future test scores. These results show that the achievement test—the standard assessment tool in the US—misses many of the skills that affect meaningful outcomes.

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

	Ninth-Grade Measure					
Outcome	Explore Test	GPA	Credits	Absences	Discipline	All
ACT Score (Grade 11)	0.78	0.22	0.05	0.10	0.02	0.79
GPA (Grade 11)	0.21	0.49	0.28	0.20	0.05	0.52
Absences (Grade 11)	0.09	0.22	0.12	0.35	0.03	0.39
Arrested within 4 Years	0.06	0.14	0.12	0.10	0.10	0.20
Grad HS within 5 Years	0.11	0.35	0.36	0.23	0.06	0.41
Enroll College within 6 Years	0.15	0.20	0.16	0.12	0.03	0.25
Grad College within 10 Years	0.17	0.17	0.07	0.09	0.01	0.23

Sources: CPS, CPD, and NSC administrative data. **Notes:** The table shows the predictive power (R^2) from a regression of the outcomes listed in the first 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. "GPA" includes the fall and spring GPAs from ninth grade. "Credits" includes two separate variables for fall and spring credits accumulated. "Absences" indicates the percentile rank of absences in 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.

Factor Model

To reduce the dimensionality of the data, we posit that two latent variables (factors) explain the variances in both the measurement system and the outcomes. We use a very similar approach for many of the analyses in this paper so we describe it here in detail.

We apply a simple three-step method. First, under the assumptions described in Section 3.2, we estimate the distribution of factors $F(\theta)$. Second, we predict factor scores $\hat{\theta}_i$

using the estimated distribution. However, using factor scores can lead to attenuation bias in the estimated coefficients due to measurement error. To account for this possibility, we adopt the "bias-avoidance" method for calculating the factor score, as described in Skrondal and Laake (2001).[30] To account for estimation error in $\hat{\theta}_i$, we calculate the standard errors by estimating 100 bootstrap samples. Third, we estimate the following equation for each outcome k:

$$Y_{ki} = \alpha_{Yk}\hat{\theta}_i + U_{Yki}.$$

We make some assumptions to improve interpretability. Given the number of measures that we use, it is possible to identify a model with more than two factors. We test this possibility and find that two factors are sufficient to explain the variation in the measures.[31] We also assume that the cognitive and non-cognitive skill factors are uncorrelated. In a sensitivity test, we allow for the possibility that the two factors are correlated and test whether the correlation is different from zero. We fail to reject the hypothesis ($p = .79$), thereby confirming our original assumption.

The factor model has advantages over the simple R^2 estimates in Table 2, which are difficult to interpret for two main reasons. First, the R^2 estimates do not correct for measurement error in the independent variables. We show that this is the case for several variables. Second, an independent variable might have a low value of R^2 even if it is important, simply because there is little variation in the variable in the sample.

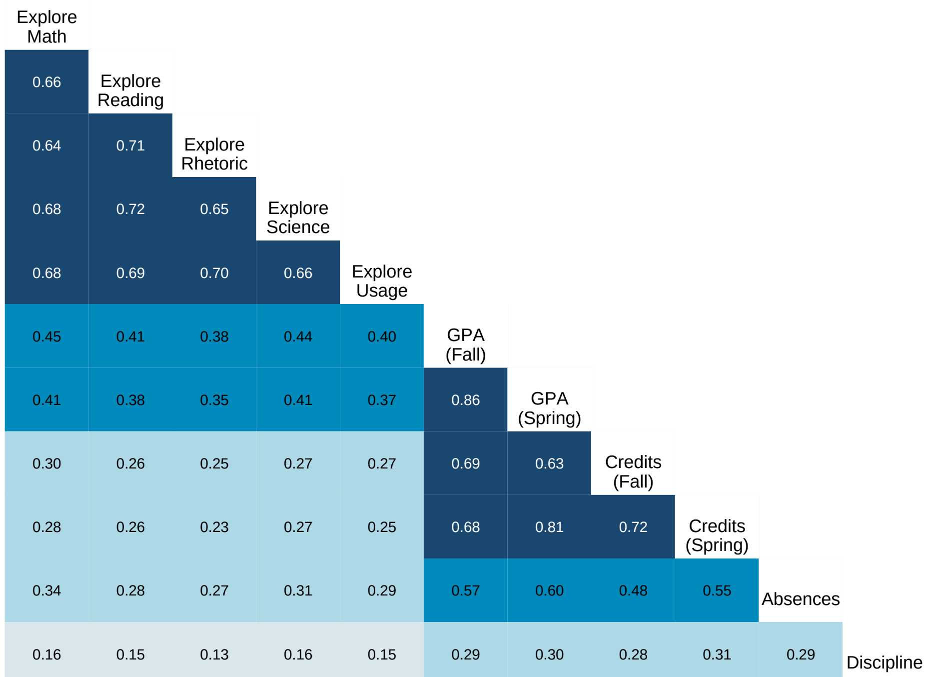
Figure 2 shows the variance explained by cognitive skills, non-cognitive skills, and measurement error for (a) each of the measures used in the system and (b) the outcomes that we analyze. We do not display the variance explained by distance to school, because it accounts for a negligible amount of the variance for all measures. This analysis confirms the patterns

[30] Table W37 in Section W9.1 of the Web Appendix shows that our main results are unchanged if we use a two-step maximum likelihood procedure rather than the factor score method.

[31] 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, supporting our decision to use two factors. See Figures W2 and W3 in Web Appendix Section W6. We determine the number of factors using the Kaiser criterion and Horn's adjustment for sampling error (Kaiser, 1960; Horn, 1965).

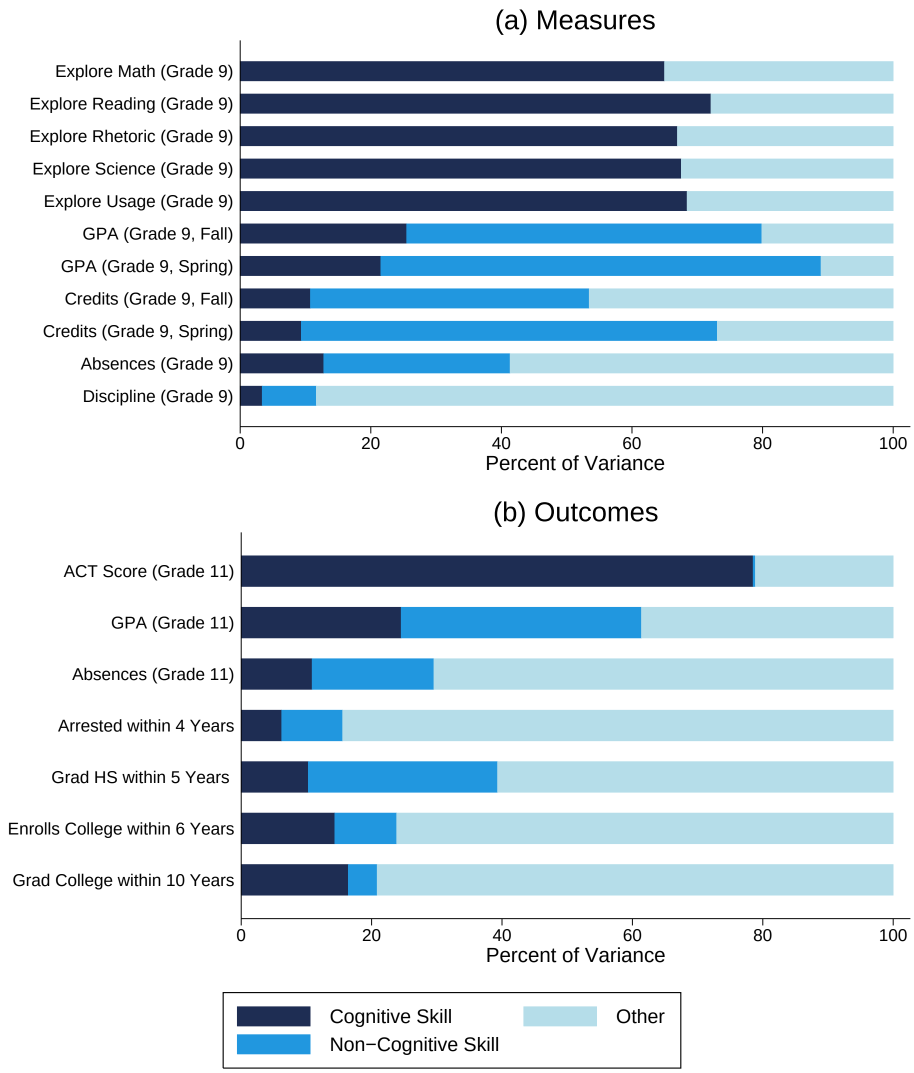
shown in Figure 1. Non-cognitive skills explain much additional variance in the outcomes beyond cognitive skills.

Figure 1 Correlations among Ninth-Grade Measures



Source: CPS administrative data. **Notes:** The figure shows the correlations among various ninth-grade measures. The sample is restricted to students who were enrolled in ninth grade for the first time during the 2003–2004 school year. All variables have been normalized so that higher values represent beneficial outcomes. "Explore Math," "Explore Reading," "Explore Rhetoric," "Explore Science," and "Explore Usage" refer to the subscores on the subtests of the Explore test given in ninth grade. "GPA (Fall)" and "GPA (Spring)" are the grade point averages from fall and spring. "Credits (Fall)" and "Credits (Spring)" are the total credits earned in the first two semesters of ninth grade. "Absences" indicates the percentile rank of absences in ninth grade. "Discipline" refers to the total disciplinary infractions from Groups 3–6 in ninth grade.

Figure 2 Variance Decomposition of the Measurement System and Various Outcomes



Sources: CPS, CPD, and NSC administrative data. **Notes:** Panel (a) shows a decomposition of the variance of each measure into a component due to cognitive skill, a component due to non-cognitive skill, and a component due to measurement error. Distance to school explains a negligible amount of the variable 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.

4.3 Description of the Final Sample

In this section we discuss how we construct our sample of students. 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 eligible for OneGoal in the sense that they were enrolled in a CPS school during the second semester of tenth grade.

There are three main limitations of our data. First, not all schools are required to report all academic indicators. Charter schools do not report absences, grades, or disciplinary infractions. Of the thirty-four schools that OneGoal has served within Chicago, twelve are charter schools. In Section 5.2 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 non-participants from OneGoal schools, and 186,707 non-participants 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. Importantly, we also use data from before OneGoal was introduced to CPS assuming no non stationarity trends. Third, we do not observe measures for students who transfer out of CPS during high school, so we remove them from the analysis once they transfer.

5 Main Results

5.1 Characteristics of OneGoal Participants

In this section we analyze the pre-program characteristics of OneGoal participants and non-participants. This analysis motivates the need to account for pre-program differences in evaluating the program.

We analyze a sample of students who could have participated in OneGoal in the sense that they were enrolled in a CPS high school during the second semester of tenth grade. We distinguish between OneGoal participants ("participants"), non-participants who attended a school that at some point offered OneGoal ("OneGoal school non-participants"), and non-participants who attended a school that never offered OneGoal ("non–OneGoal school non-participants").

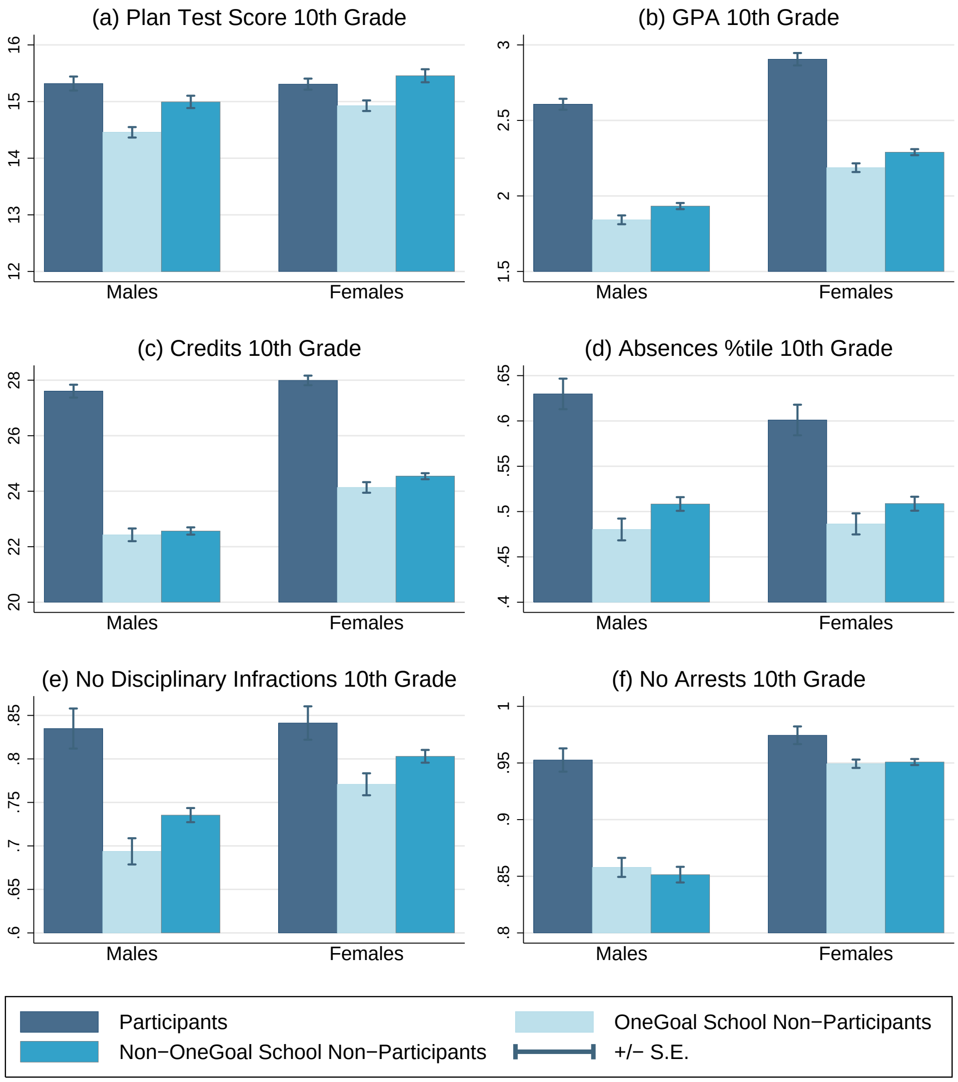
Figure 3 displays the characteristics of OneGoal participants, non-participants in OneGoal schools, and non-participants from non-OneGoal schools, sorted by gender.[32] Due to the large sample sizes, most of the differences between OneGoal participants and other students are statistically significant. However, the magnitude of the difference in test scores is small. Compared to non-participants in OneGoal schools, participants score between 0.3 and 0.8 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 one-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 two-thirds less likely to be arrested during tenth grade compared to non-participants. Both male and female participants have higher GPAs, complete more credits, and have far fewer absences.

These patterns suggest that OneGoal participants have higher non-cognitive skills than

[32] For tables by the cohort, see Tables W15–W28 in Web Appendix Section W7.

Figure 3 Pre-Program Characteristics of OneGoal Participants and Non-Participants (Tenth Grade)



Sources: OneGoal, CPS, and CPD administrative data. **Notes:** The graphs show the average tenth-grade measures for OneGoal participants, non-participants in OneGoal schools, and non-participants in other schools. The vertical lines ("⊢⊣") represent the standard errors for each mean. All variables have been normalized so that higher values represent beneficial outcomes. "Plan 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 were calculated using 100 bootstrap samples and allow for clustering at the school-cohort level.

their peers in CPS. We summarize these differences by applying the factor structure described in Section 4.2. Figure 4 shows the distribution of extracted cognitive and non-cognitive skill factor scores for OneGoal participants, non-participants from OneGoal schools, and non-participants from other schools. The scores are standardized to have a mean of zero and a standard deviation of one for the full CPS sample. The distribution of cognitive skills is similar for OneGoal participants and non-participants. In contrast, the distribution of non-cognitive skills for OneGoal participants is narrower and shifted far to the right compared to those of non-participants, suggesting that OneGoal selects students with higher non-cognitive skills.[33] Accounting for these pre-program differences is vital for estimating the treatment effects.

5.2 Estimated Treatment Effects from Matching

In this section, we estimate the effects of OneGoal using a model in which we compare OneGoal participants to other students by matching on unobserved skills and other pre-program characteristics. 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.

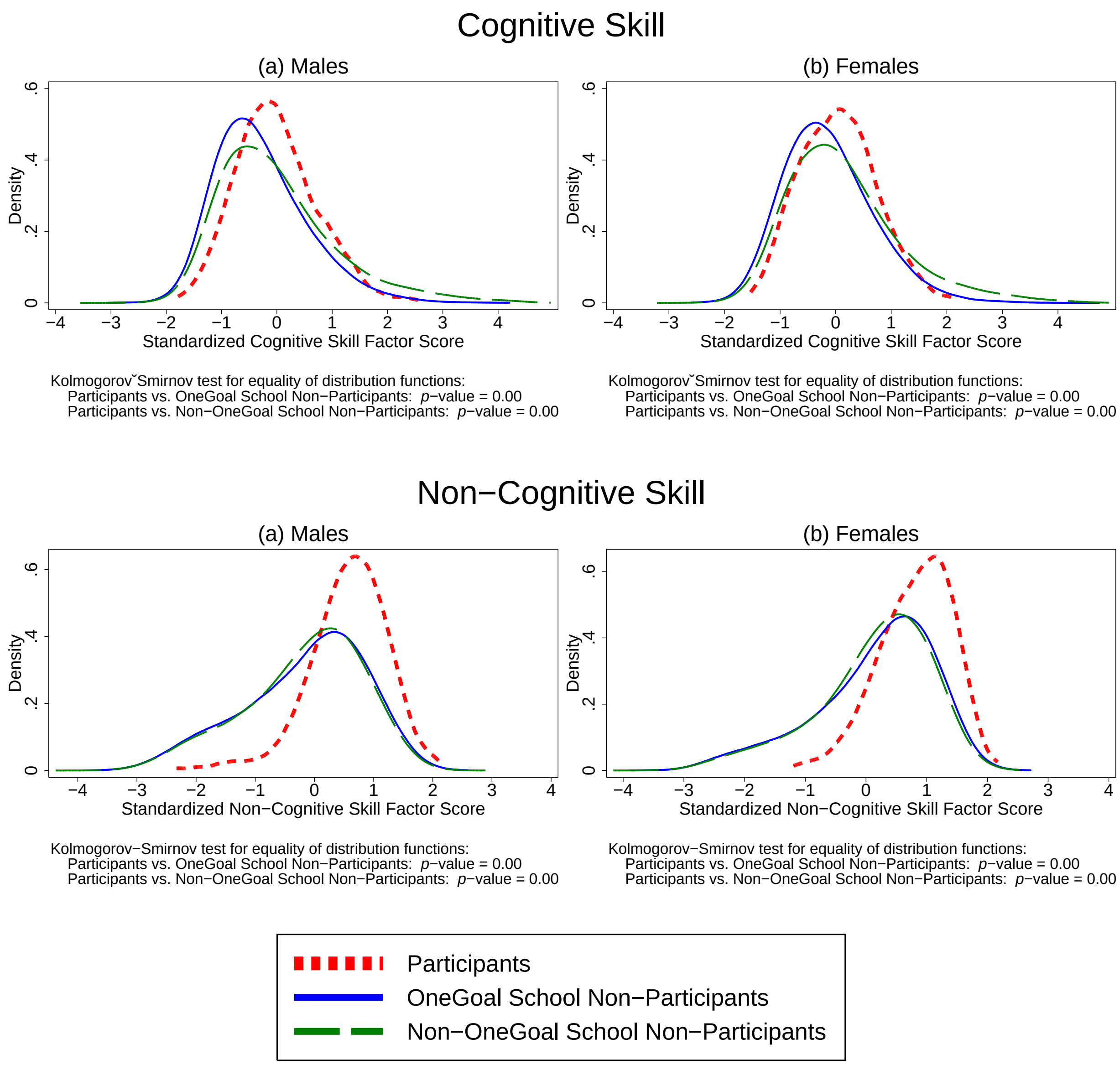
In principle, we could conduct our analysis for the entire sample of schools in CPS. We restrict the sample to schools that at some point offer OneGoal in order to create control groups consisting of students who are most similar to OneGoal participants. In a sensitivity check, we use the full sample and find very similar estimates. We define treatment as whether a student was accepted into OneGoal, not whether they completed the full program to account for the potential for selective attrition out of OneGoal.[34]

We apply the simple three-step procedure described in Section 4.2 to estimate the following

[33] These trends are also apparent when considering the distribution of the individual measures. See Tables W29–W31 in Section W7 of the Web Appendix.

[34] 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. See Figure W14 in Web Appendix Section W9. The estimates are similar if we drop these students. See Figure W14 in Web Appendix Section W9.

Figure 4 Distribution of Cognitive and Non-Cognitive Skills for OneGoal Participants and Non-Participants



Sources: OneGoal and CPS administrative data. **Notes:** 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 non-cognitive 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 non-cognitive measures are also allowed to depend on the cognitive measures. The scores have been standardized to have a mean of zero and a standard deviation of one using a sample of students who were first-time tenth graders in CPS between 2005 and 2013.

equation for each outcome k:

$$Y_{ki} = \beta_{Yk} X_i + \alpha_{Yk} \hat{\theta}_i + \delta_k A_i + U_{Yki},$$

where A_i is an indicator for whether person i was accepted into OneGoal, X_i are basic demographic characteristics, and $\hat{\theta}_i$ is the predicted factor score. We calculate standard errors by estimating 100 bootstrap samples and allow for errors to be clustered at the school-cohort level—the level at which eligibility for OneGoal varies in our sample. We also estimate a non-linear version using a two-step maximum likelihood procedure, as well as a non-parametric version that uses propensity score matching. These other specifications lead to similar estimates.[35]

Figure 5 presents results based on matching on latent cognitive and non-cognitive skills for males and females.[36] For each gender, the first bar shows the difference after only controlling for basic demographics between OneGoal participants and non-participants, the second bar shows the effect after additionally controlling for cognitive skills, and the third bar shows the results after additionally controlling for non-cognitive skills. The vertical lines ("⊢⊣") represent the standard errors for each mean, and the symbols on the bars indicate the results from tests of significance. Years are measured relative to when students would have started OneGoal had they been recruited into the program.

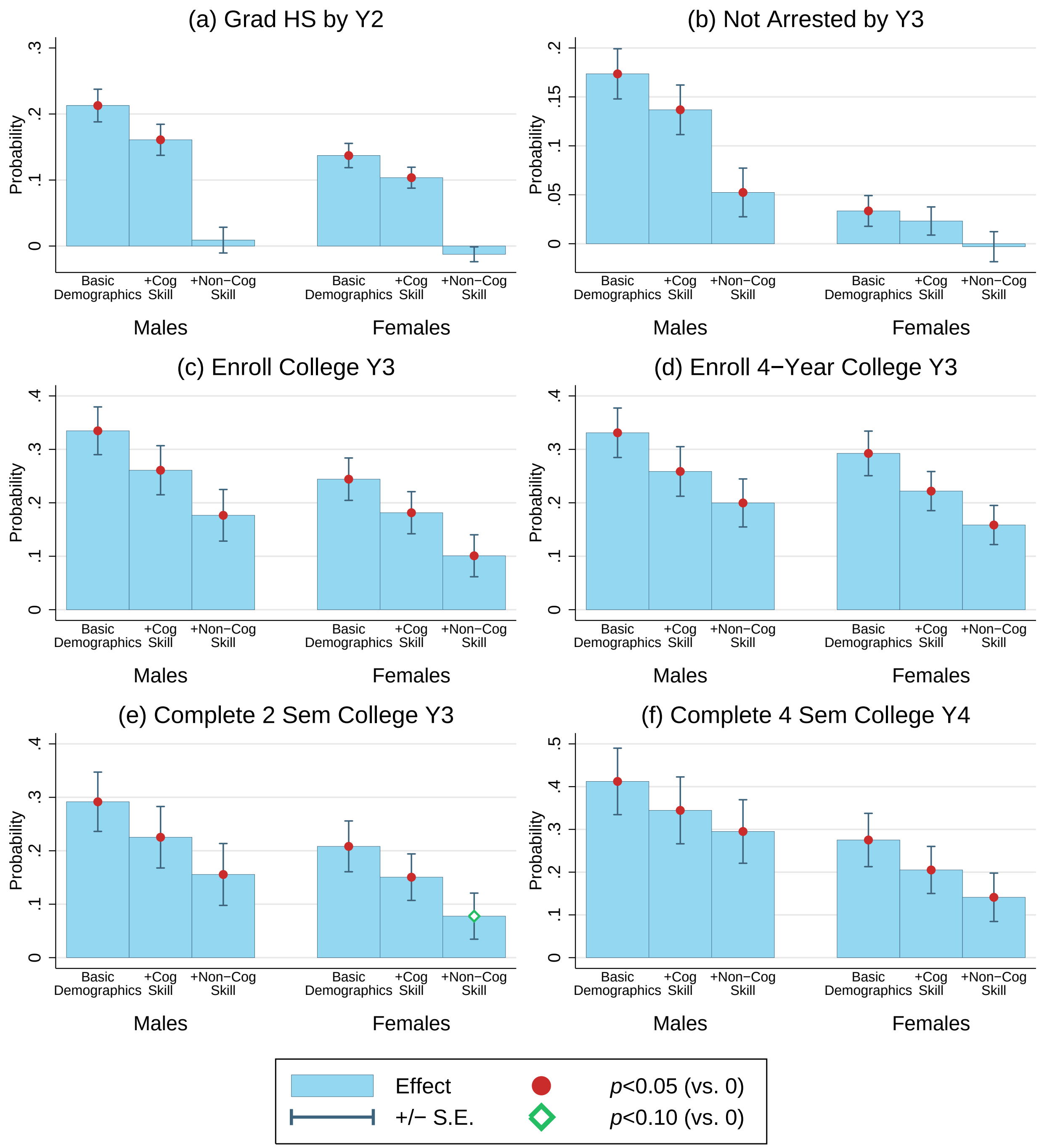
The figure reveals three striking results. First, OneGoal improves college outcomes across the board. It has the strongest effect on four-year college enrollment and persistence. Second, OneGoal has greater effects for males than for females. OneGoal improves arrest rates for males, but not for females, and it has a stronger impact on college outcomes for males. Third, accounting for non-cognitive skills is important. If we only controlled for demographics and cognitive skill, we would have estimated that OneGoal increases high school graduation by

[35] Table W37 in Section W9.1 of the Web Appendix shows that our main results are unchanged if we use a two-step maximum likelihood procedure rather than the factor score method. Table W45 in Section W9.5 of the Web Appendix shows that our main results are unchanged if we use propensity score matching.

[36] Table W32 in Section W8.1 of the Web Appendix shows the same results for both genders combined.

10–15 percentage points for both males and females. After controlling for non-cognitive 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 non-cognitive skills.

Figure 5 Treatment Effects for Main Outcomes Based on Matching



Sources: OneGoal, CPS, CPD, and NSC administrative data. **Notes:** The figure shows the effects of OneGoal for each outcome listed on the top of each panel. The labels along the x-axis indicate the control variables included. The vertical lines ("⊢⊣") 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 non-cognitive 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 non-cognitive measures are also allowed to depend on the cognitive measures. The standard errors were calculated using 100 bootstrap samples and allow for clustering at the school-cohort level.

5.3 Mediation Analysis

This section considers whether OneGoal operates by improving skills as opposed to 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 the pre-program cognitive and non-cognitive 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 non-cognitive skill.

Table 3 shows the effects of OneGoal on eleventh-grade academic indicators based on the methodology described in Section 5.2. The estimates are adjusted for basic demographics, cognitive skill, and non-cognitive skill and are normalized so that positive numbers correspond to beneficial outcomes. For males, OneGoal has a significant effect on ACT scores, absences, credits, and GPAs. For females, OneGoal does not have an effect on ACT scores but does improve absences, credits, and GPAs. These findings suggest that OneGoal might work in part because it improves both cognitive and non-cognitive skills.

To investigate this further, we consider how OneGoal affects our measures of cognitive and non-cognitive skills in eleventh grade and how changes in skills are associated with improvements in our main outcomes. We decompose the effect of OneGoal on outcomes into its effect on cognitive skills, non-cognitive skills, and other factors. Let θ_C^0 and θ_N^0 be cognitive and non-cognitive skills in tenth grade, before students are recruited into OneGoal, and let θ_C^1 and θ_N^1 be cognitive and non-cognitive skills in eleventh grade, the first year in which students are eligible to participate.

Following recent studies in economics that model skill formation, we allow past skills to affect future skills.[38] We also allow for the possibility that OneGoal participation could

[37] We do not have access to a CPS measure of cognitive skill in twelfth grade so we focus on eleventh grade.
[38] See, for example, Cunha and Heckman (2007, 2008) and Cunha et al. (2010).

Table 3 Estimated OneGoal Effects on Eleventh Grade Academic Indicators after Adjusting for Basic Demographics, Cognitive Skill, and Non-Cognitive Skill

Outcome	Males	Females
ACT Score	0.50***	0.01
	(0.13)	(0.11)
Absences %tile	0.05***	0.03**
	(0.02)	(0.02)
Discipline	0.08	-0.06
	(0.06)	(0.06)
GPA	0.11***	0.13***
	(0.04)	(0.04)
Credits	1.27***	0.91**
	(0.44)	(0.49)

Sources: OneGoal and CPS administrative data. **Notes:** The table shows the effects of OneGoal for each outcome listed in the left column after adjusting for basic demographics, cognitive skill, and non-cognitive skill. The outcomes are normalized so that higher values represent beneficial outcomes. The standard errors were calculated using 100 bootstrap samples and allow for clustering at the school-cohort level. * 10% significance; ** 5% significance; *** 1% significance.

improve skills:

$$\theta_C^1 = \gamma_{C0} + \gamma_{C1}\theta_C^0 + \gamma_{C2}\theta_N^0 + \phi_C A_i + \eta_{Ci},$$
$$\theta_N^1 = \gamma_{N0} + \gamma_{N1}\theta_C^0 + \gamma_{N2}\theta_N^0 + \phi_N A_i + \eta_{Ni},$$

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

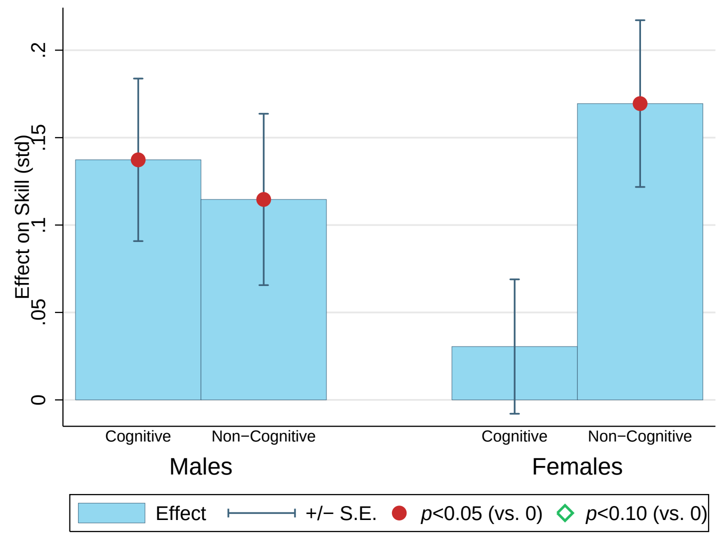
$$Y_{ki} = \beta_{Yk} X_i + \alpha_{Yk}\theta^1 + \delta_k A_i + U_{Yi}.$$

The total effect of OneGoal is decomposed as follows:

$$\text{Total Effect} = \underbrace{\alpha_{Yk}\phi}_{\text{Indirect Effect}} + \underbrace{\delta_k,}_{\text{Effect through Other Skills or Information}}$$

where $\phi = [\phi_C, \phi_N]$.We estimate the model using a two-stage maximum likelihood approach

Figure 6 Effect of OneGoal on Eleventh-Grade Cognitive and Non-Cognitive Skills



Sources: OneGoal and CPS administrative data. **Notes:** The figures show the effect of OneGoal on the indicated skill. The skills have been normalized to have a variance of one separately for each gender. The vertical lines("⊢⊣") represent the standard errors for each mean, and the symbols on the bars indicate the results from tests of significance. The standard errors were calculated using 100 bootstrap samples and allow for clustering at the school-cohort level.

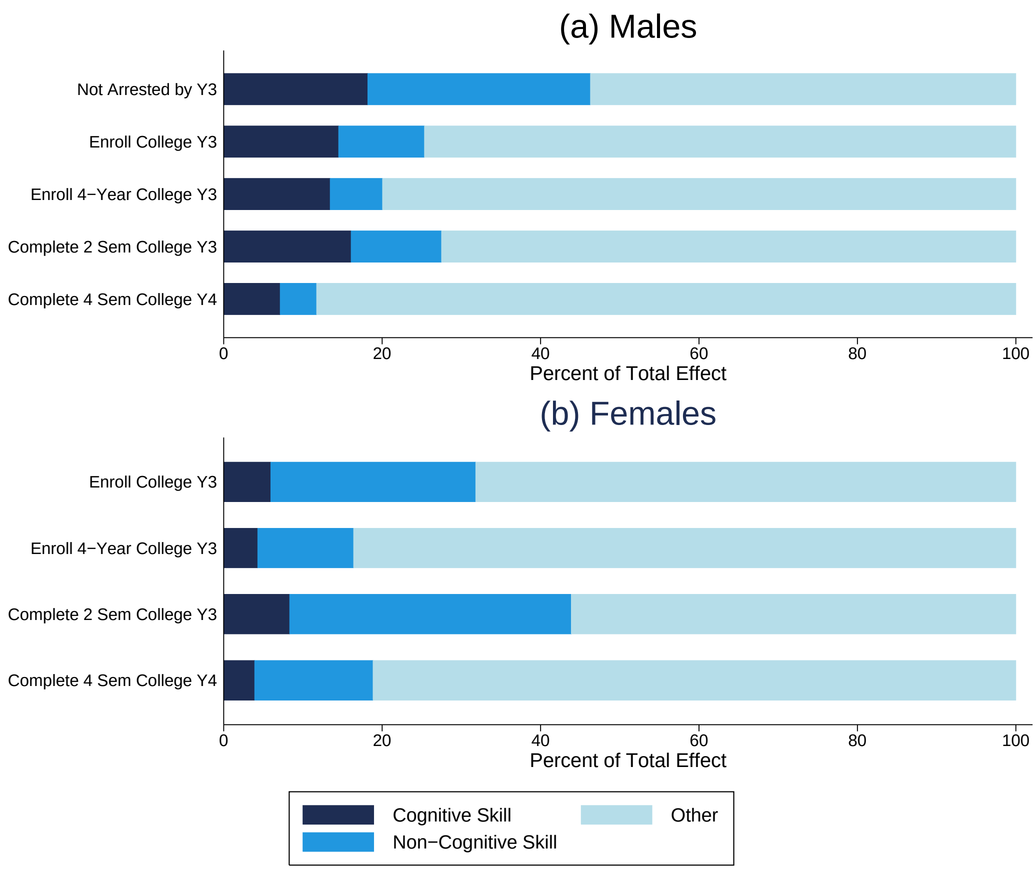
and calculate the standard errors using 100 bootstrap draws.

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

Figure 7 shows the percent of the total effect that can be attributed to improvements in cognitive skill, non-cognitive skill, or other factors. We display only the outcomes for which we estimate a statistically significant effect in the analysis presented in Figure 5. For males, improvements in both cognitive and non-cognitive 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 effect. This result is consistent with Figure 6, which shows that OneGoal had little impact 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. Consistent with this interpretation, the effect due to other factors is similar to the estimates from the Bettinger et al. (2012) intervention which only provided information and assistance in applying for financial aid (see Section 2.2). These estimates suggest, however, that providing mentorship and skill development also plays an important role.

Figure 7 Percent of Total Effect due to Cognitive Skill, Non-Cognitive Skill, and Other Factors



Sources: OneGoal, CPS, CPD, and NSC administrative data. **Notes:** The figure shows the percent of the total effect that can be attributed to improvements in cognitive skill, improvements in non-cognitive skill, or other factors. Only outcomes with effects that are statistically different from zero are displayed.

5.4 Treatment Effect Heterogeneity

In this section we study whether OneGoal has different benefits for students starting at different skill levels. To study this issue, we estimate a version of the model described in Section 5.2 in which we allow skills to interact with OneGoal participation indicator:

$$Y_{ki} = \beta_{Yk} X_i + \alpha_{Yk}\hat{\theta}_i + \delta_k A_i + \gamma_k \hat{\theta}_i A_i + U_{Yki}.$$

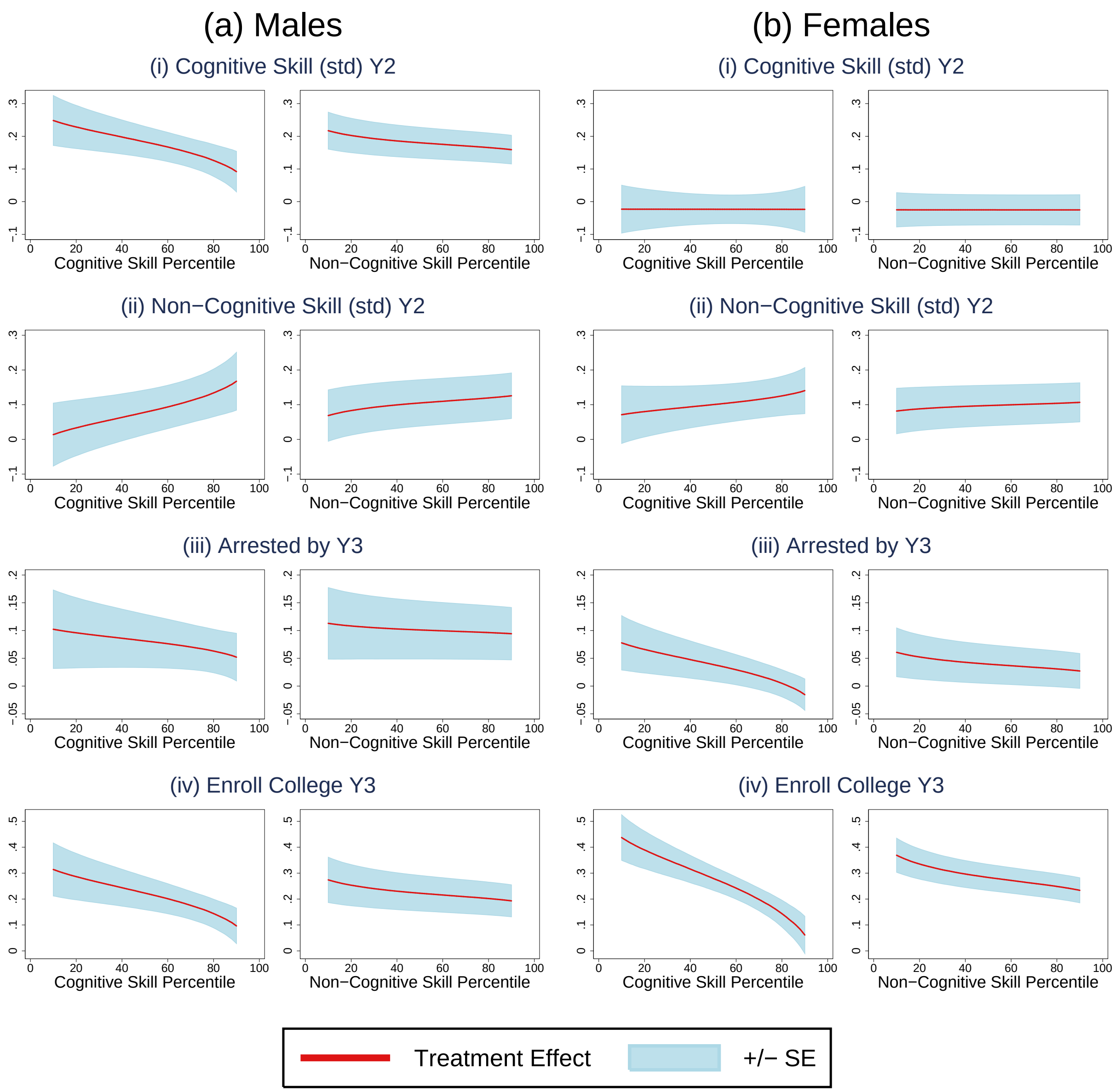
Figure 8 displays the results from this analysis, separated by gender. For each outcome, the left panel shows the treatment effect as a function of pre-program cognitive skill and the right panel shows the effect as a function of pre-program non-cognitive skill. For each panel, the other skill is fixed at the median. The band represents plus and minus one standard error of the estimate. The range on the x-axis indicates the percentiles of skill for the treated group.[39]

Overall, the estimates suggest that students with the lowest pre-program skills benefit the most from the intervention consistent with the literature. The only exception is that males with the highest cognitive skill see the biggest improvements in non-cognitive skill. These results suggest that OneGoal might be most effective if it targeted students with lower levels of both cognitive and non-cognitive skills.

The findings in this section might also explain why males benefit more than females. Male participants tend to have lower levels of non-cognitive skills than their female counterparts (see Figure 4 and Figure 3 in Section 5.1.)

[39] Specifically, we limit the skills to range from the tenth to the ninetieth percentile within the treated group.

Figure 8 Effect of OneGoal on Eleventh-Grade Cognitive and Non-Cognitive Skills



Sources: OneGoal, CPS, CPD, and NSC administrative data. **Notes:** The figures show the effect of OneGoal on the indicated outcome by level of pre-program cognitive skill (left panel of each pair) and pre-program non-cognitive skill (right panel of each pair). The band represents plus and minus one standard error of the estimate. All variables have been normalized so that higher values represent beneficial outcomes. The cognitive skill factor score is based on the subscores from the reading, English rhetoric, English usage, science, algebra, and geometry subtests of the Plan test. The non-cognitive skill factor score is 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 non-cognitive measures are also allowed to depend on the cognitive factor. Standard errors are calculated using 100 bootstrap samples.

5.5 Estimated Treatment Effects Based on Using OneGoal Eligibility as an Instrument

This section applies a difference-in-difference instrumental variable method. It exploits the fact that OneGoal was introduced into different schools at different times, so that some cohorts of students were ineligible to participate in OneGoal simply because it was not offered in their school. We use eligibility as an instrumental variable to compare students who were eligible to participate with those who were not, similar to a randomized experiment. In order 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.

This method yields estimates that are less precise than our matching method, so we take several steps to increase the precision. In our matching estimates, 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 not 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.[40]

We could use the full sample of schools, including those that never adopted OneGoal at all. This would help increase the precision in the estimates of cohort effects or any other covariates common across schools. When we use the full sample, we find similar estimates for arrests and college enrollment, but the estimates for OneGoal's effect on high school

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

graduation are implausibly high (around 30 percentage points) and statistically significant. We investigate the reason for this difference and find that some non-OneGoal schools have negative time trends. In some cases the schools shut down during our study window. In one case, OneGoal had approved a school for participation but the principal did not implement OneGoal because the school was expected to close. For this reason, we restrict our main analysis to schools that offer OneGoal at some point.

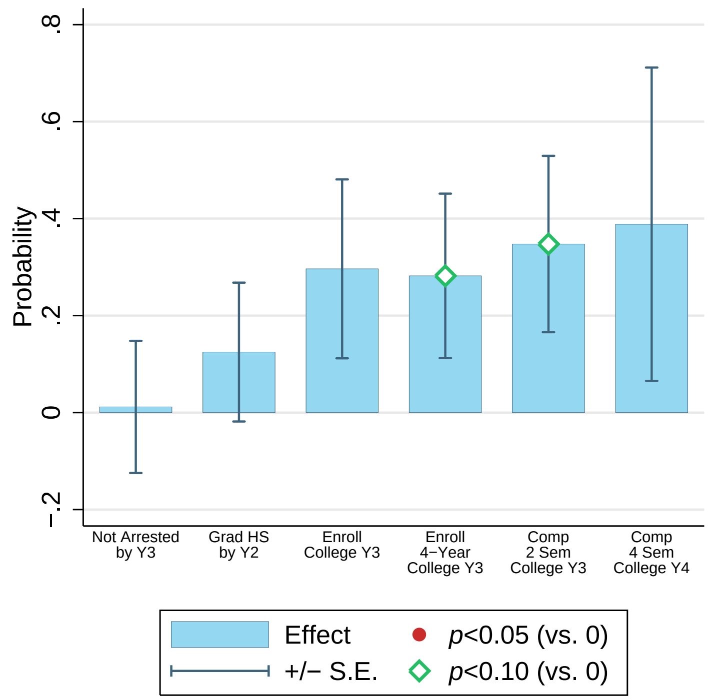
We use two-stage least squares to estimate the first and second stage equations:

$$A_{ics} = \beta^0 X_{ics} + \delta^0 Z_{ics} + f_c^0 + f_s^0 + \varepsilon_{ics}^0,$$
$$Y_{ics} = \beta X_{ics} + \delta A_{ics} + f_c + f_s + \varepsilon_{ics},$$

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 is eligible for OneGoal, f_c is a fixed effect for the cohort, and f_s is a fixed effect for the school. We allow for errors to be clustered at the school-cohort level. In our main specification, X_{ics} includes only gender.

Figure 9 shows empirical results from this analysis. As with the matching analysis, we find that OneGoal has a strong effect on college outcomes but no statistically significant effect on high school graduation. In a sensitivity check, we include the other covariates to account for possible changes within 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. The results of this analysis are broadly consistent with our matching approach, but they are much less precise so we place less weight on them. We are reassured to find similar results using a very different source of variation.

Figure 9 Treatment Effects for Main Outcomes Based on Using OneGoal Eligibility as an Instrument



Sources: OneGoal, CPS, CPD, and NSC administrative data. **Notes:** The figure shows the effects of OneGoal for each outcome listed along the x-axis. The vertical lines("⊢⊣") represent the standard errors for each mean, and the symbols on the bars indicate the results from tests of significance. The standard errors were calculated using 100 bootstrap samples and allow for clustering at the school-cohort level. The model allows for school and cohort fixed effects and includes a dummy variable for gender.

5.6 College Persistence

As discussed in Section 4.3, 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 non-cognitive skills. For this reason, our matching analysis is limited to the first two years of college, because the sample sizes are too small to conduct the matching analysis. When we include charter schools, we have a large sample size that allows us to consider college enrollment through the third year of college, but we cannot control for pre-program skills.

We study this additional data to understand whether there are any "fade-out" effects of OneGoal. Figure 10 displays the unadjusted college enrollment rates for OneGoal participants and non-participants by semester for three years after students would have graduated from high school.[41] The figure suggests that the raw differences are roughly constant over time, suggesting that the effects presented in Figure 5 likely persist at least through the third year of college.

6 Sensitivity Checks

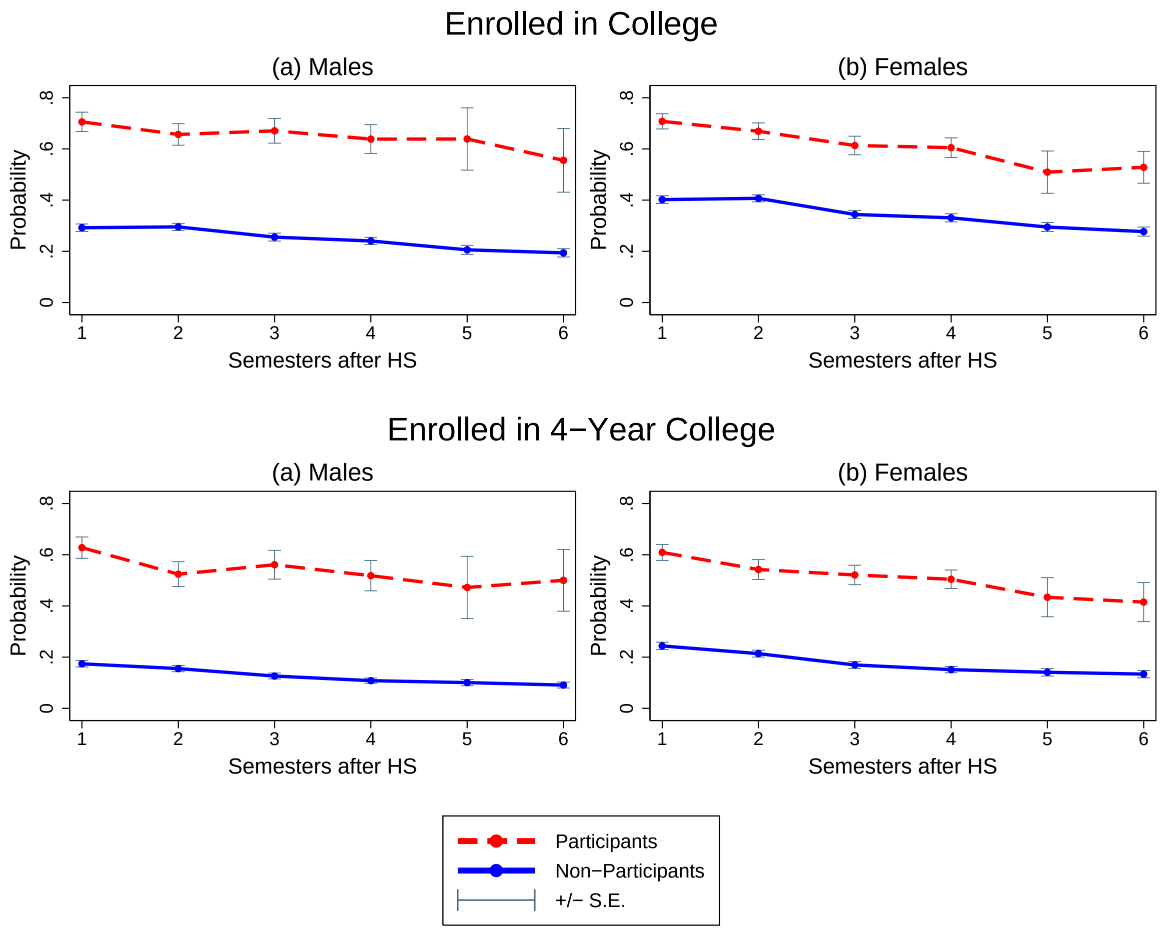
In this section we summarize results from a number of sensitivity checks. Web Appendix Section W9 contains the full set of results.

6.1 Using Teacher Assessments and Denied Applicants

In this section, we further validate our measures of skills using the scores from assessments used to rate OneGoal applicants during their interviews. The scores are each measured on a scale of 1 to 5 and are ratings of the "five leadership principles" (ambition, integrity, professionalism, resilience, and resourcefulness). For five cohorts, we have these assessments

[41] Note that the estimate of college enrollment is lower than what is posted on OneGoal's website because OneGoal bases their report on data from the NSC and by confirming enrollment through OneGoal Program Directors and registrars for colleges that do not report to the NSC, whereas our table reports estimates based only on the NSC.

Figure 10 College Persistence



Sources: OneGoal, CPS, and NSC administrative data. **Notes:** The figure shows the unadjusted college enrollment rates for OneGoal participants and eligible non-participants in OneGoal schools. The vertical lines ("⊢⊣") represent the standard errors for each mean. The standard errors allow for clustering at the school-cohort level. "Semesters after HS" refers to the number of semesters that have passed since a student would have graduated from high school if they had graduated on time.

for both denied applicants ($N = 75$) and accepted applicants ($N = 100$).

For these students, we consider how incorporating the scores from the assessments would affect our estimates of the OneGoal treatment effects. Table 4 shows the OneGoal treatment effects for eleventh-grade outcomes for the cohorts for which we observe accepted and denied applicants. The outcomes have been normalized so that positive values are beneficial. The first column shows the effects when controlling for only basic demographics and the measures of cognitive and non-cognitive skills. The second column shows the effect when additionally controlling for the rubric scores. The estimates for these outcomes are consistent with those presented for other samples. OneGoal has an effect on absences, credits earned, and arrests, although the sample size might be too small to draw strong conclusions on all of the outcomes. This analysis also accounts for the possibility that the students who choose to apply to OneGoal might be more motivated, because the control group also applied to OneGoal.

Table 4 Estimated OneGoal Effects When Controlling for Rubric Scores and CPS Measures among Applicants

	Mean	**Model**
Outcome	(0)	(1)
GPA Year 1	-0.12	-0.12
	(0.10)	(0.12)
Absences %tile Year 1	0.04	0.09**
	(0.04)	(0.05)
ACT Score	-0.15	-0.12
	(0.36)	(0.41)
Credits Year 1	3.09*	4.69***
	(1.34)	(1.51)
Discipline Year 1	-0.01	0.05
	(0.14)	(0.16)
Number of Arrests Year 1	0.10***	0.10***
	(0.03)	(0.03)
Rubric Scores	○	●
CPS Measures	●	●

Sources: OneGoal, CPS, and CPD administrative data. **Notes:** The table shows the effects of OneGoal for each outcome listed in the left column. The filled circles at the bottom of the table indicate the controls used in the model. "Rubric Scores" is the sum of the Ambition, Integrity, Professionalism, Resilience, and Resourcefulness teacher ratings of leadership. "CPS Measures" include race, gender, a predicted cognitive factor score, and a predicted non-cognitive factor score. The cognitive factor score is based on the subscores from the reading, English rhetoric, English usage, science, algebra, and geometry subtests of the Plan test. The non-cognitive factor score is 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 non-cognitive measures are also allowed to depend on the cognitive measures. * 10% significance; ** 5% significance; *** 1% significance.

7 Conclusion

We evaluate OneGoal, a program that attempts to help disadvantaged high school students complete college, in part by helping them apply to college and in part by improving non-cognitive skills that are not captured by test scores. It teaches non-cognitive 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 females. Our panel is too short to estimate its effect on college graduation. Improvements in cognitive and non-cognitive skill account for about one-third of these effects. We find that OneGoal also improves outcomes through another factor, possibly the information which participants receive about applying to college. These results suggest that programs that combine targeted information with skill development are promising.

In conducting the evaluation, we devise a way to measure non-cognitive skills using administrative data available in most schools. This measure outperforms test scores in predicting arrests and high school graduation.

This evaluation demonstrates the importance of accounting for non-cognitive skills. First, we show that before they enter the program OneGoal participants tend to have higher levels non-cognitive skills than non-participants. If we did not account for these differences, we would overestimate the effects of OneGoal. Second, we find that OneGoal had a relatively small effect on test scores (cognitive skills) but that it had large effects on other outcomes, such as college enrollment. If we had only measured test scores and not non-cognitive 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 test scores to assess students and schools.

References

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

ACT, Inc. (2013a). *The ACT Explore Technical Manual.* Iowa City, IA: ACT, Inc.

ACT, Inc. (2013b). *The ACT Plan Technical Manual.* Iowa City, IA: ACT, Inc.

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

Allensworth, E. M. and J. Q. Easton (2007). What matters for staying on-track and graduating in Chicago Public High Schools: A close look at course grades, failures, and attendance in the freshman year. Technical report, Consortium on Chicago School Research at the University of Chicago.

Almlund, M., A. Duckworth, J. J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education*, Volume 4, pp. 1–181. Amsterdam: Elsevier.

Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 5*, pp. 111–150. Berkeley: University of California Press.

Aos, S., R. Lieb, J. Mayfield, M. Miller, and A. Pennucci (2004). Benefits and costs of prevention and early intervention programs for youth. Report 04-07-3901, Washington State Institute for Public Policy, Olympia, WA.

Benda, B. B. (2005). The robustness of self-control in relation to form of delinquency. *Youth & Society 36*(4), 418–444.

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. *Quarterly Journal of Economics 127*(3), 1205–1242.

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. Last accessed online February 11, 2013. http://www.mdrc.org/sites/default/files/full_491.pdf.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008, Fall). The economics and psychology of personality traits. *Journal of 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: Special Issue on Personality and Economics), 315–320.

Carrell, S. E. and B. Sacerdote (2013). Late interventions matter too: The case of college coaching in New Hampshire. Working Paper 19031, NBER.

Chicago Public Schools (2002, June). Adopt Uniform Discipline Code for Chicago Public Schools for 2002–2003 school year. Board Report 02-0626-PO01, Chicago, IL.

Chicago Public Schools (2003, July). Adopt Uniform Discipline Code for Chicago Public Schools for 2003–2004 school year. Board Report 03-0723-PO02, Chicago, IL.

Chicago Public Schools (2004, July). Adopt Uniform Discipline Code for Chicago Public Schools for 2004–2005 school year. Board Report 04-0728-PO2, Chicago, IL.

Chicago Public Schools (2005, July). Adopt Uniform Discipline Code: The student code of conduct for all Chicago Public Schools students for the 2005–2006 school year. Board Report 05-0727-PO1, Chicago, IL.

Chicago Public Schools (2006a, August). Adopt Student Code of Conduct (formerly the Uniform Discipline Code) for Chicago Public Schools effective September 12, 2006. Board Report 06-0823-PO3, Chicago, IL.

Chicago Public Schools (2006b, July). Adopt Student Code of Conduct (formerly the Uniform Discipline Code) for Chicago Public Schools for 2006–2007 school year. Board Report 06-0726-PO8, Chicago, IL.

Chicago Public Schools (2007, July). Adopt Student Code of Conduct for Chicago Public Schools for the 2007–2008 school year. Board Report 07-0627-PO5, Chicago, IL.

Chicago Public Schools (2008, July). Adopt Student Code of Conduct for Chicago Public Schools for the 2008–2009 school year. Board Report 08-0723-PO5, Chicago, IL.

Chicago Public Schools (2009, July). Adopt Student Code of Conduct for Chicago Public Schools for the 2009–2010 school year. Board Report 09-0722-PO5, Chicago, IL.

Chicago Public Schools (2010, July). Adopt Student Code of Conduct for Chicago Public Schools for the 2010–2011 school year. Board Report 10-0728-PO2, Chicago, IL.

Chicago Public Schools (2011, July). Adopt Student Code of Conduct for Chicago Public Schools for the 2011–2012 school year. Board Report 11-0727-PO2, Chicago, IL.

Chicago Public Schools (2012, June). Adopt Student Code of Conduct for Chicago Public Schools for the 2012–2013 school year. Board Report 12-0627-PO1, Chicago, IL.

Chicago Public Schools (2013, August). Adopt a Grade Point Average policy. Board Report 13-0828-PO6, Chicago, IL.

Cook, P. J., K. Dodge, G. Farkas, R. G. Fryer, Jr, J. Guryan, J. Ludwig, S. Mayer, H. Pollack, and L. Steinberg (2014, January). The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth: Results from a randomized experiment in Chicago. Working Paper 19862, NBER.

CORE (2013). ESEA flexibility request for window 3. Report, California Office to Reform Education.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008, Fall). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Duckworth, A. (2012, November). Reference bias. Presented at the University of Chicago.

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.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. Chang, and S. Grantham-McGregor (2013, June). Labor market returns to early childhood stimulation: A 20-year followup to an experimental intervention in Jamaica. Working Paper 19185, NBER.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2012). Education, health and wages. Technical report.

Heckman, J. J. and T. Kautz (2012, August). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464. Adam Smith Lecture.

Heckman, J. J. and T. Kautz (2014). Fostering and measuring skills: Interventions that improve character and cognition. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, pp. 341–430. Chicago, IL: University of Chicago Press.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010, February). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics 94*(1–2), 114–128.

Heckman, J. J. and S. Navarro-Lozano (2004, February). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics 86*(1), 30–57.

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 70, pp. 4779–4874. Amsterdam: North-Holland.

Heckman, J. J. and E. J. Vytlacil (2007b). 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 J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 71, pp. 4875–5143. Amsterdam: North-Holland.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika 30*(2), 179–185.

Jackson, C. K. (2013). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina. Working Paper 18624, NBER.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement 20*(1), 141–151.

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. *OECD*. Forthcoming.

Kemple, J. and J. C. Snipes (2000, March). Career Academies: Impacts on student engagement and performance in high school. Report, MDRC. Last accessed on September 26, 2014. http://www.mdrc.org/sites/default/files/full_45.pdf.

Kemple, J. and C. Willner (2008, June). Career Academies: long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood. Report, MDRC. Last accessed on September 26, 2014. http://www.mdrc.org/sites/default/files/full_50.pdf.

Millenky, M., D. Bloom, and C. Dillon (2010, May). Making the transition: Interim results of the National Guard Youth ChalleNGe evaluation. Report, MDRC. Last accessed online September 26, 2014. http://www.mdrc.org/sites/default/files/full_434.pdf.

Millenky, M., D. Bloom, S. Muller-Ravett, and J. Broadus (2011, June). Staying on course: Three-year results of the National Guard Youth ChalleNGe evaluation. Report, MDRC. Last accessed online September 26, 2014. 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), 931–964.

Reynolds, A. J., J. A. Temple, S.-R. Ou, I. A. Arteaga, and B. A. B. White (2011, July). School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups. *Science 333*(6040), 360–364.

Roder, A. and M. Elliot (2011). A promising start: Year-Up's initial impacts on low-income young adults' careers. Technical report, Economic Mobility Corporation.

Roder, A. and M. Elliot (2014). Sustained gains: Year-Up's continued impact on young adults' earnings. Technical report, Economic Mobility Corporation.

Rodríguez-Planas, N. (2012, October). Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the United States. *American Economic Journal: Applied Economics 4*(4), 121–139.

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. *Journal of Cross-Cultural Psychology 38*(2), 173–212.

Schochet, P. Z., J. Burghardt, and S. McConnell (2008, December). Does Job Corps work? Impact findings from the National Job Corps Study. *American Economic Review 98*(5), 1864–1886.

Skrondal, A. and P. Laake (2001, December). Regression among factor scores. *Psychometrika 66*(4), 563–576.

Tierney, J. P., J. Baldwin Grossman, and N. L. Resch (1995). Making a difference: An impact study of Big Brothers Big Sisters. Report, Public/Private Ventures.

Tyler, R. W. (1940, September). The place of evaluation in modern education. *The Elementary School Journal 41*(1), 19–27.

Williams, B. (2012). Identification of factor models. Unpublished dissertation, University of Chicago, Department of Economics.