The Effects of Letters of Recommendation in the Youth Labor Market*

Sara B. Heller (University of Michigan & NBER)†
Judd B. Kessler (University of Pennsylvania & NBER)

December 9, 2021

Abstract

Youth employment has been near historic lows in recent years, and racial gaps persist. This paper tests whether information frictions limit young people's labor market success with a field experiment involving over 43,000 youth in New York City. We build software that allows employers to quickly and easily produce letters of recommendation for randomly selected youth who worked under their supervision during a summer youth employment program. We then send these letters to nearly 9,000 youth over two years. Being sent a letter generates a 3 percentage point (4.5 percent) increase in employment the following year, with both employment and earnings increases persisting over the two-year follow-up period. By posting our own job advertisement, we document that while treatment youth do use the letters in applications, there is no evidence of other supply-side responses (i.e., no increased job search, motivation, or confidence); effects appear to be driven by the demand side. Labor market benefits accrue primarily to racial and ethnic minorities, suggesting frictions may contribute to racial employment gaps. But improved employment may also hamper on-time high school graduation. Additional evidence indicates that letters help improve job match quality. Results suggest that expanding the availability of credible signals about young workers—particularly for those not on the margin of graduating high school—could improve the efficiency of the youth labor market.

*This work was funded by the Social Policy Research Initiative at J-PAL North America. We thank the New York City Department of Youth and Community Development, the New York State Department of Labor, and the New York City Department of Education for sharing the data with us. Thanks to Charlie Brown, Alicia Modestino, Mike Mueller-Smith, Alex Rees-Jones, Ana Reynoso, Basit Zafar, and the JIM group at Princeton for helpful comments. We are particularly grateful to Julia Breitman at DYCD for all of her help along the way, to Ben Cosman at DOE for all his support, and to Alex Hirsch and Lauren Shaw for phenomenal project management and research assistance. All views in the paper are those of the authors and do not represent the positions of any data provider or government agency.

†Corresponding author: sbheller@umich.edu

1 Introduction

Youth employment lags farther behind adult employment than can be explained by school engagement, and it recovers more slowly in the wake of major shocks. Even a decade after the Great Recession, youth employment rates in summer—when teenagers are most likely work—were still hovering near their sixty-year low (DeSilver 2021). And the Covid-19 pandemic disproportionately harmed youth labor market prospects at the outset (Inanc 2020; Kochhar and Barroso 2020). Employment gaps are particularly bleak for minority youth. Unemployment and disconnection rates are 30–80 percent higher for Black and Hispanic youth than for their White peers (Bureau of Labor Statistics 2020, 2021; Spievack and Sick 2019), a pattern of disproportionate labor market involvement that is not new (Sum et al. 2014). That youth, and minority youth in particular, have such difficulty securing work is troubling given the range of evidence suggesting that labor force attachment during adolescence and young adulthood may shape employment and wage trajectories for decades (Baum and Ruhm 2016; Kahn 2010; Neumark 2002; Oreopoulos, Wachter, and Heisz 2012).

Half a century of active labor market programs have tried to address these issues, often with relatively costly efforts to improve human capital or provide job search assistance. Despite some success in developing countries and in U.S. sector-focused training (Card, Kluve, and Weber 2018; Katz et al. 2020; Crépon and Van Den Berg 2016), the frequency with which training programs fail to improve youth employment in the U.S. raises the possibility that important frictions may limit young people's access to the labor market. Their short or nonexistent work histories may leave little way to credibly signal future productivity. Employers may engage in statistical or animus-driven discrimination based on age, class, or race, or they may view participation in training programs as a negative signal itself. Alternatively, on the supply side, young people may lack the social networks, knowledge, or confidence to convert their experience into better employment outcomes.

In this paper, we explore the potential role of information frictions in the youth labor market by testing an intervention designed to mitigate them. We partner with the New York City Summer Youth Employment Program (NYC SYEP), which employs city youth to work over the summer, to run a large-scale field experiment with over 43,000 youth. We provide a random subset of youth participants with letters of recommendation from their SYEP supervisors. To make letter production on this scale feasible, we invite program supervisors to complete a survey tool, developed by our research team, that automatically turns their survey responses on individual participants into full-text letters of recommendation. When supervisors agree to produce a letter of recommendation and provide high enough ratings of a worker, that treatment youth receives a digital copy and five hard copies of the letter.

We follow treatment and control youth in administrative data for at least two years after letter distribution to measure their labor market and educational outcomes. Across a pilot after the summer of 2016 and a full-scale study after the summer of 2017, a total of 43,409 SYEP participants are in our main study sample.[1] We measure employment and earnings using unemployment insurance data from the New York State Department of Labor. We observe educational outcomes, which could be directly affected by the letters or indirectly via changes in labor force involvement, using data from the NYC Department of Education.

Partnering with the NYC SYEP provides an ideal environment to assess the role of frictions in the youth labor market. SYEPs are large social programs that provide paid work to youth—often low-income and minority youth—during the summer months. For about half of these youth, SYEP participation is their first experience in the labor market. Consequently, SYEP participants are representative of the groups likely to face barriers in their attempts to capitalize on early work experience. Indeed, while prior literature has found that SYEPs improve important outcomes including criminal justice involvement and mortality, multiple randomized controlled trials suggest they do not have consistently positive average effects on future employment (Davis and Heller 2020; Gelber, Isen, and Kessler 2016; Modestino 2019).

Our results suggest a sizable impact of the letter of recommendation intervention. We find that being sent a letter increases the likelihood that a young person is employed by over 3 percentage points in the year after receiving the letter, a 4.5 percent increase relative to the 70 percent of their control group counterparts who work.[2] Employment effects persist over time, with impacts remaining positive and statistically significant over the cumulative two-year follow-up period; youth who are sent a letter are 2 percentage points more likely to have a job over the next two years, a 2.3 percent increase relative to the control complier mean of 84 percentage points. Cumulative earnings are at least 4.4 percent higher for those sent the letter ($p=0.10$), with different adjustments for skewness suggesting considerably bigger and more statistically significant effects (between 10 and 20 percent, $p<0.05$).

That simply providing a few pieces of paper improves employment and earnings suggests an important role for information frictions in the youth labor market. Our treatment could be mitigating frictions on either the supply side or the demand side. On the supply side, letters may give youth information about what makes them valuable to employers and the confidence to apply to jobs; on the demand side, letters may give employers a clearer signal

1. Our empirical strategy involves stacking panels for the two cohorts, so youth can appear in the data more than once. In total, we have 43,409 observations on 41,633 unique individuals.

2. This effect is 250% as large as prior estimates of the effect of the summer program itself on employment. Gelber, Isen, and Kessler (2016) finds that the NYC SYEP increased employment by 1.2 percentage points in the post-program year by encouraging youth to participate in SYEP again.

about the abilities of a particular youth or make a youth's application more salient.

To assess the mechanism driving our results, we ran an additional data-collection exercise to measure job-seeking behavior among a subset of our sample. We invited 4,000 participants from both treatment and control groups in the 2017 cohort to apply for a short-term online job working for us. Youth in our treatment group were no more likely to apply for our job and no more likely to check a box asking to be considered for a more-selective, higher-paying opportunity, suggesting that our employment effects are not being driven by increased motivation, job search, or confidence.[3] That there was no detectable difference in application behavior among treatment and control youth suggests that the letters work by changing how employers view applicants, rather than how applicants behave.

The only behavioral difference was in the use of the letter itself: Treatment youth were 267% more likely to submit a letter of recommendation as part of their application (4.5 percent of control applicants and 16.5 percent of treatment applicants included a letter in their application). Given the lack of other supply-side responses, it seems possible that this difference in letter usage is the key driver of outcome changes, i.e., that the letters work *only* when employers actually see them. If so, we might view the treatment-control difference in letter use as an implied first stage for the letters' effects on compliers who use the letter in job applications. Back-of-the-envelope calculations—which involve scaling our employment estimates by the implied first stage to approximate the relevant LATE—suggest that actually using the letter is associated with up to a 15 percent increase in employment in the first year and about $1,400 in additional earnings (also about 15 percent) over two years.

In the presence of demand-side frictions, finding additional ways to communicate credible information about youth applicants to potential employers may help youth succeed in the labor market. Consistent with this idea, we find that employment and earnings increases are concentrated among youth who are more highly rated by their SYEP employers, suggesting an important role for information transmission in the letters. We also find that the labor market benefits accrue primarily to minority youth, despite the fact that they receive letters with lower average ratings than White youth. This result suggests that minority youth may face larger frictions in the youth labor market.

One might worry that letters could lead employers to *incorrectly* update beliefs about candidates. If employers' priors about who obtains letters lead them to believe applicants will be higher productivity employees than they actually are, short-term increases in employment could represent bad matches and be followed by increased churn. Because we can track how

3. While it is possible that the increased outside employment among treatment youth affected the decision to apply to our job, application rates did not differ either between treatment and control youth who were employed elsewhere during the quarter we solicited applications, or between treatment and control youth who were not employed elsewhere.

many consecutive quarters employees work at the same employer, we can test this question directly. We find no evidence of increased turnover; treatment youth work more quarters at the same number of jobs, such that some job spells get longer. This set of findings suggests the letters are actually helping workers and employers make more successful job matches.

We collect information on educational outcomes in addition to employment data for two reasons. First, letters could have a direct educational effect if shown to teachers or guidance counselors: other work has shown that teachers' and other adults' beliefs about young people directly affect their outcomes, even when the information that changed those beliefs is fictitious (Rosenthal and Jacobson 1968; Bertrand and Duflo 2017). Letters could also help with college applications if young people have few other sources of recommendations. Second, working during high school could pull young people out of school. There is a general consensus in the current literature that working a small amount has little effect on schooling but that working more than 20 hours is harmful (Buscha et al. 2012; Staff, Schulenberg, and Bachman 2010; Monahan, Lee, and Steinberg 2011; Baum and Ruhm 2016; Ruhm 1997). However, the lack of exogenous variation in this literature means that it is still unclear whether a shock to employment would have a causal effect on school success, at least outside of a setting that mandates continued school enrollment as part of offering a term-time job (Le Barbanchon, Ubfal, and Araya 2020).

For the nearly 20,000 youth in our study who we can observe in New York City public high schools, we find few significant changes in educational performance. But among the subset of youth for whom we can observe graduation, we find that letters of recommendation slow down—but do not appear to stop—high school graduation by pulling people into the labor market. For the average high school student, the welfare implications of efforts to reduce information frictions in the labor market depends on how future career trajectories are affected by the value of additional work experience, and how those benefits compare to the cost associated with a longer time spent in high school. In the meantime, given the heterogeneity we document, policy efforts to provide employers with credible information on minority high school students may minimize the risk of substitution away from school by targeting those not on the margin of graduating on time.

While our employment effects are large, they are consistent with past research that suggests providing even a small amount of information about job-seekers can be quite powerful. In response to fictitious applications in audit studies (Agan and Starr 2018; Kaas and Manger 2012) and to the suppression of information in the labor market (Bartik and Nelson 2019; Doleac and Hansen 2020), employers show less discrimination when they have more information about candidates. Providing performance information has also been shown to increase short-term employment in two different kinds of labor markets. Pallais (2014) randomly hires nearly 1,000 workers on an online labor market platform (oDesk), provides randomly selected workers with more-detailed public performance reviews, and finds that workers with no prior experience benefit on the platform from being hired and rated, while those with prior work experience benefit from the detailed reviews over the next two months. Abel, Burger, and Piraino (2020) find that encouraging a subset of 1,300 job seekers in South Africa to secure letters of recommendation increases job search and survey measures of being employed among women, but not overall, after three months.

We build on this prior work by exploring the impact of letters of recommendation among 43,000 young people in a large, urban U.S. labor market. The setting expands the study of information frictions to an environment where, unlike oDesk or South Africa, employers can potentially access a range of richer signals about youth applicants (e.g., more widespread, visible employment histories or knowledge about local high schools and GPAs) but also face higher hiring costs (e.g., a minimum wage or more burdensome paperwork). Using administrative data, we can observe employment and earnings at jobs across New York State for two years after treatment, as well possible spillovers on high school outcomes for the study youth who are also enrolled in New York City high schools. Our data also allow us to explore specific questions unanswered by the prior literature, such as whether short-term increases in employment are simply a (temporary) result of encouraging bad matches.

Our study provides new evidence that information frictions do prevent young people, especially non-Whites, from securing successful employment. Given that employers seem to value credible information about applicants, finding additional ways to provide personalized information could help improve labor market outcomes among low-income, minority populations like those in our study. Such interventions might be best targeted at those who are not on the margin of graduating on time to avoid harming educational attainment.

The impact of scaling up efforts to facilitate letters of recommendation (or other credible signals) will depend on general equilibrium effects that we cannot directly measure within our study. Welfare effects of expanding letter distribution in general equilibrium could go in either direction. It is possible that reducing information frictions could increase overall employment by helping employers fill vacancies they would otherwise have left open in the face of too much uncertainty. It is also possible that youth with recommendation letters may simply displace those without them (although this is unlikely to have happened within the context of our control group, given that there are about one million 15- to 24-year-olds in the NYC labor market and we sent fewer than 9,000 letters across two years). Even the welfare implications of full displacement are not obvious, however, since policymakers may value the distributional changes or efficiency gains from better matches, even if there were no net change in employment.

Additional research on exactly how letters change employers' decision-making processes would help to predict the welfare consequences of broader efforts to facilitate credible productivity signals. For now, this study provides new evidence on the role of information frictions in discouraging employment and earnings for young people. Such factors potentially limit the impact of programs designed to improve their skills and future labor market outcomes. Fortunately, it may not be particularly costly to reduce these frictions by communicating to employers about applicants' strengths.

2 Setting, Experiment, and Data

2.1 Setting

We partner with the New York City Summer Youth Employment Program (NYC SYEP), implementing our experiment with youth who participated in the summer of 2016 or the summer of 2017. The NYC SYEP is administered by the NYC Department of Youth and Community Development (DYCD). Since a post-Great Recession minimum enrollment of 29,416 youth, enrollment grew steadily to nearly 70,000 youth in 2017. In our program years, the NYC SYEP provided youth with six weeks of paid work during July and August. All NYC residents aged 14–24 were eligible to apply for the SYEP program, though 40% of eventual participants were aged 16–17. Participants in the program were provided with jobs with private sector (45%), non-profit (41%), and public sector (14%) employers. The NYC SYEP directly pays youth for their work with their matched employers at the New York State minimum wage ($11.00/hour in 2017). Youth payroll totaled $83 million in 2017, or roughly $1,200 per youth participant. The NYC SYEP had a total program cost of $127 million in 2017. Over 80% of this cost was funded by the City of New York, with a majority of its remaining funding coming from New York State (see *SYEP Annual Summary* 2017).

2.2 Letter of Recommendation Experiment

We received SYEP data from DYCD on a subset of participants from the 2016 NYC SYEP (n=16,478) and all of the participants in the 2017 NYC SYEP (n=66,763). The program data identified each youth's summer work site and the supervisor or supervisors for the youth at that work site. Using these data, we limited our sample in several ways. First, since we needed to contact supervisors to ask them to complete the letter of recommendation survey, we excluded youth supervised by someone without an email address in the data. Second, we excluded some youth at large work sites to avoid making the survey unmanageable for a single supervisor. In particular, if any supervisor was linked to more than 30 treatment youth, then we randomly selected 30 treatment youth to be included in the survey. We applied the same restriction for the control youth in the survey.[4] In total, this left a sample of 69,222 SYEP participants who were included on at least one survey. Figure 1 traces through this and the subsequent steps of how youth moved through the study.

To generate recommendation letters, we built a survey tool that sent a personalized survey to each supervisor asking about the youth who they supervised that summer (i.e., the youth linked to them in the DYCD data).[5] The email inviting each supervisor to participate explained the letter of recommendation program, included a link to the personalized survey tool, and encouraged them to participate (a sample of the email from 2017 is shown as Appendix Figure A.1). Supervisors were given approximately two weeks to complete the survey, and we sent up to two reminder emails to supervisors who had not yet completed it. For the 2016 cohort, we emailed 3,297 supervisors at the end of September (initial emails went out on 09/29/16). For the 2017 cohort, we emailed 11,877 supervisors in October (initial emails went out on 10/12/17).

The survey began with a brief explanation for supervisors that if they rated a youth positively enough, their responses to the survey questions might be used to construct letters of recommendation. A link to an example letter was provided to aid in the explanation. Respondents were then asked to confirm that they had been a SYEP supervisor during the preceding summer (see screens at the start of the survey in Appendix Figure A.2). Once a respondent confirmed being a supervisor, they were shown the list of treatment youth linked to them in the DYCD data, listed alphabetically by last name.[6] Supervisors selected which youth they had directly supervised and were asked a set of questions about each youth (supervisors were asked about the youth they selected in random order). The survey asked supervisors for an overall rating on the youth's performance and whether they would be willing to answer questions that would turn into a letter of recommendation for the youth (see Figure A.2 for screenshots of the survey). If they were willing, they were also invited

4. To ensure that neither the treatment nor control group exceeded the 30-person-per-survey limit, we randomly assigned treatment and control status prior to making these sample restrictions. Since youth were randomly selected to be excluded, random assignment is still only a function of random variables.

5. The data did not link every youth to a single supervisor. Sometimes, multiple supervisors were listed for a single work site, such that it was not clear which youth reported to which supervisor or if a youth reported to multiple supervisors; in these cases, we assumed the latter for the purposes of constructing our survey tool. Consequently, youth could be listed on more than one survey. Sometimes, a single supervisor was listed for multiple work sites. If the names of the work sites suggested they might be connected (e.g., multiple branches of the same store), we treated them as one work site for the purposes of constructing the survey tool. In the survey, we asked supervisors to confirm the youth that worked for them and to provide the names of others who might have supervised youth so we could include them in the letter of recommendation program as well. If more than one supervisor rated a young person, we generated the letter from the survey with the highest rating, breaking ties by prioritizing letters that included employer contact information, and then those with the most positive responses about the youth.

6. Note that confirming one's identity and position as an SYEP supervisor is how we count "starting" the survey, a definition that is relevant below.

to include their contact information on the letter of recommendation to serve as a reference (97 percent of eventual letters included contact information). They then rated the youth on several attributes, shown in Figure 2.

After the supervisors answered questions about treatment youth, they were asked one question each about control youth—the same question about the overall rating on the youth's performance—all on one screen (see Appendix Figure A.3). They were told that these youth would not be included in the letter of recommendation program.

A total of 5,854 supervisors (39 percent of all supervisors we emailed) opened the survey and confirmed that they had supervised SYEP youth during the preceding summer. In total, 43,409 young people were on a started survey, 29,887 (69 percent) of whom were given an overall rating by employers.

The software we built for this project converted the supervisors' survey responses on treatment youth into formatted letters of recommendation populated with sentences for each youth attribute. For each positively rated attribute, the letter included a dynamically constructed sentence. For example, if in response to the question "How was $< youth\ name >$ at communicating?" the supervisor selected "Very effective," a sentence would appear in the letter that read: "$< Youth\ name >$ was a very effective communicator." Whereas, if the supervisor selected "Not effective" or "Somewhat effective" in response to that question, the sentence about communication would not be included in the letter.

We assigned each attribute to a potential paragraph. If the supervisor rated the youth positively enough on enough attributes to construct a particular paragraph, the paragraph was included in the letter. As long as two paragraphs could be included, the letter was generated for the youth. This procedure ensured that any letters of recommendation our survey tool generated had enough positive things to say about the youth to provide a positive letter that would not be too sparse. Our software produced letters of recommendation as PDFs on official DYCD letterhead. The letters ended with "Sincerely," followed by the name of the supervisor and work site. A short note in the footer of the letter described our letter of recommendation pilot program. Figure 3 shows a sample letter.

In total, we generated and sent 8,780 letters (1,805 in 2016 and 6,975 in 2017). We uploaded digital copies of these letters to Dropbox with a link sent to the youth for whom emails were known (1,737 in 2016 and 6,720 in 2017).[7] In addition, we mailed five physical copies of the letters via USPS to each youth along with a cover letter providing context and suggested uses for the letter (see Appendix Figure A.4 for a sample cover letter; similar

7. About 56 percent of letter recipients clicked the link in their email to view the letter digitally. Many SYEP youth create an email solely for the purpose of the online application and then abandon it, so some letter recipients may not have seen the email containing the link to the digital copy of the letter.

text was sent to youth via email along with the link to the soft copy of the letter).[8] All letters of recommendation were sent in time for winter holiday hiring in the year after SYEP participation (letters were sent to youth in early-December 2016 for the 2016 cohort and in mid-November 2017 for the 2017 cohort).

2.3 Job Application Data

To understand the mechanisms through which letters of recommendation might impact labor market outcomes of treatment youth, we advertised a job to a subset of the youth in our data, solicited job applications, and hired youth ourselves. We composed a job listing for a one-time, remote, paid work assignment, emailed the job listing to 4,000 randomly selected subjects from our 2017 cohort, and observed their job application behavior. The sample was evenly split among treatment and control youth from the letter of recommendation experiment (i.e., youth who had been eligible and ineligible to receive the letter of recommendation) who also had an email address in the data so we could send them the job application.

The job was described as being with a professor at the University of Pennsylvania who was looking for former NYC summer job participants for a short-term and flexible job. The job description highlighted several qualifications: "responsible," "self-motivated," having an "enthusiastic approach," and offered compensation of $15/hour. A link to an application with a deadline to submit an application was included at the bottom of the job description (see the email invitation sent to youth with the job description in Appendix Figure A.5).

Youth who clicked the link in the email were taken to a job application that asked a few standard contact, background, and employment experience questions. Our application also provided an optional space to upload up to three "supporting documents (e.g. resume or other documents that might strengthen your application)." The application did not explicitly mention uploading letters of recommendation, but it would have been easy for youth to upload the soft copy of the letter of recommendation provided to them in our experiment (see the screenshot of this prompt in Appendix Figure A.6).[9] This upload interface allowed us to measure whether youth provided supporting materials—including a letter of recommendation—with their applications and to assess whether this differed by treatment and control youth.

Finally, to assess the confidence of youth in our study, we gave applicants the opportunity

8. Of the 8,780 sets of letters mailed to youth, 127 were returned as undeliverable.

9. We intentionally avoided explicitly mentioning a letter of recommendation to see if youth in our study would choose to upload a letter without a specific prompt to do so. We saw this as realistic to job applications in practice where a youth could choose to provide a potential employer with a letter of recommendation even if one was not specifically requested.

to check a box on the application to be considered for a more selective, higher-paying position ($18/hour) that required a stronger application. The application told them explicitly that being considered for the more selective position would not affect their chances at being selected for the regular job.

All those who submitted an application that included their name, email address, and at least 1 additional field were hired.[10] The job itself was an online survey of multiple choice questions. These questions asked youth about their experiences job-seeking and considering college, as well as about their career and education goals. At the end of the survey, there were free-response questions about the youth's experience in SYEP.[11] Workers were instructed to finish everything they could within a two-hour time frame. All youth who initiated the job-task (n=227) were paid for two hours of work via a mailed, pre-loaded debit card (so our job does not appear in the administrative data on employment and earnings).

2.4 SYEP Administrative Data

Administrative data from the NYC SYEP comes from the NYC DYCD, which runs the program. We received data on a subset of participants of the 2016 NYC SYEP and all participants of the 2017 NYC SYEP. The data on SYEP participants include identifiers (e.g., name, date of birth, and social security number) that allow us to match to various data sources; demographics (e.g., gender, race, and pre-SYEP education status) that allow us to test for balance across treatment and control; and contact information (e.g., mailing address and email address) that we used to send letters of recommendation to treatment youth. We define racial/ethnic categories based on the self-reported categories in the application, making the classifications mutually exclusive (e.g., "White" only captures non-Hispanic Whites). We also received information on the work site where the youth worked for the summer and information about the supervisors at that work site, including name and email address. We use the information on work site and supervisor to send the letter of recommendation surveys.

2.5 NYS Department of Labor Data

We obtained earnings and employment data from the New York State Department of Labor (NYSDOL). Data came from NYSDOL's quarterly Unemployment Insurance (UI) dataset, which covers formal sector employment, excluding self-employment or farming income. The

10. To ensure our hiring for the more selective job was incentive compatible with our instructions about higher selectivity, the youth needed to click the box asking to be considered and needed to complete one or more of the open response questions in addition to fulfilling the requirements for the standard job.

11. Youth hired for the more selective job were asked additional open response questions that required more thoughtful consideration.

data include employer name, employer FEIN, employer address, employer NAICS, and amount paid in each quarter. NYSDOL analysts matched SYEP participants to UI data using social security number. When multiple profiles in the NYSDOL data shared the same social security number, we used name to disambiguate the UI data. In total, 99 percent of SYEP youth in our letter of recommendation experiment were matched to the NYSDOL data with no difference between treatment and control youth ($\beta = 0.001, p = 0.128$).[12]

We use data from Q1 (January–March) of 2010 through Q4 (October–December) of 2019. This window provided considerable baseline data as well as two years of outcome data after letters were sent to SYEP participants in our treatment group for each study cohort.[13]

2.6 NYC Department of Education Data

Education data come from the NYC Department of Education (DOE).[14] The DOE used name, date of birth, and gender to perform a probabilistic match between our study sample and their records between the 2015–2016 and 2019–2020 school years, inclusive. SYEP applicants fail to match because they never appear in the DOE system (e.g., always attended private school), matched to more than one student record (DOE treats multiple matches on the same name and birth date as a non-match), or because typographical errors or name changes prevented identifying a study participant's education records. Overall, 88 percent of our main sample matched to a DOE student record, with no treatment-control difference in match rates ($\beta = -0.003, p = 0.359$). Within the sample that matched to a DOE student record, 7,643 had no active enrollment within our 2015–2020 data. These students were largely old enough to have left school prior to 2015 (their average age at randomization is 19.7), although some may have transferred to private or non-NYC districts prior to the start of our data. This leaves 69.9% of our main sample with at least some education information in the data, with no treatment-control difference ($\beta = -0.003, p = 0.436$).

12. In theory, everyone in our data should have matched to the data, since they were all listed as a SYEP participant during the summer prior to the program. Some of the non-workers may not have matched to the UI data despite having worked due to typographical mistakes or incorrect SSNs. Others may not have ever been paid by SYEP despite being listed as a participant in their data, and so not actually have received any wages to be reported to the UI system.

13. Letters were sent in Q4 (October–December) of 2016 or 2017, depending on cohort. Consequently, we have additional quarters of data for the youth in the 2016 cohort, but we limit the analysis to the period we can observe for full years for both cohorts, and we stop prior to any influence from Covid-19.

14. At the request of the data provider, when we merge DOE data with the rest of our study data, we exclude the self-reported citizenship status that appears on the SYEP application, so that education outcomes are never linked to citizenship status. SYEP application data also provides spotty information on whether youth live in public housing or are on public assistance; those are also never linked to DOE data.

3 Method of Analysis

This section discusses how we perform the analysis in this paper. In Section 3.1, we describe our sample definitions and our outcomes of interest for each data source. In Section 3.2, we describe our empirical approach, including our regression specifications. In all sections, we note cases where we deviated from our pre-analysis plan with accompanying explanations for these choices.[15]

3.1 Sample Definitions and Outcomes

3.1.1 Labor Market Sample

Our main sample to explore labor market outcomes consists of the 43,409 SYEP participants who were on a survey that a SYEP supervisor started (i.e., the SYEP participant appeared on at least one survey in which the supervisor clicked the link inviting them to take the survey and confirmed on the first page of the survey—prior to viewing which youth were on the survey or what their treatment status was—that they supervised youth that summer). This excludes the 25,813 youth who were randomized and placed on a survey that no supervisor ever opened.

We pre-specified this subsample of youth on a started survey as a key sample of interest, because neither treatment nor control youth on *unopened* surveys could have actually received treatment. This kind of non-compliance mechanically reduces statistical power and is orthogonal to treatment status, so we focus on the subsample with a first stage of 0.404 (rather than the first stage of 0.254 when we include youth on unopened surveys).[16] As a result, the treatment effect of receiving a letter of recommendation in our main sample is representative of the population of youth whose supervisors both had an up-to-date email address in the DYCD data and were willing to click on an invitation to participate in the letter of recommendation program. The estimates from this sample of youth might differ from the treatment effect on the broader sample of all SYEP youth, because different types of youth are placed into jobs with different types of supervisors.[17]

15. The pre-analysis plan can be found at https://osf.io/8zwdr/

16. While we pre-specified this subsample as a key sample of interest, our main sample included all SYEP participants that we randomized, because we did not anticipate that only 39% of supervisors would open the survey and that such a large fraction (i.e., over one third) of the sample would be on an unopened survey. For completeness, we present and discuss results for this larger sample in Appendix Section A.8. We choose to emphasize the results from our smaller sample in the main text, because the power gains from focusing on this subsample give better insight into the effect of the letter of recommendations on the sample of youth who might actually have been eligible to be treated, given the actions of their supervisors.

17. Appendix Section A.8 shows that youth who were on unopened surveys are indeed observably different than the youth in our control group of opened surveys on demographics and employment outcomes, although not in their likelihood of applying to our job posting. As such, it is possible that forcing supervisors to rate

Since supervisor non-response was driven by an inability to reach supervisors by email or by a lack of supervisor interest or capacity to complete the survey, limiting our analysis to this sample does not interfere with the integrity of random assignment (i.e., until the supervisors reached the substantive survey questions, they had no way of knowing which youth would be included in the survey or which youth would be in the treatment or control groups). As discussed below, Table 1 shows that our main sample is balanced across treatment and control youth.

3.1.2 Labor Market Outcomes

We pre-specified a primary focus on annual earnings, winsorized to deal with outliers, along with alternative methods to adjust for skewness as robustness checks. We pre-specified an indicator for any employment as a secondary outcome. Our main analysis shows employment and earnings in the first year after randomization (4 quarters including the quarter the letters were sent), the second year, and cumulatively, winsorized at the 99.5^{th} percentile, as well as log(earnings+1). Since the +1 transformation effectively manufactures the proportional change from zero, in Appendix Section A.1.1 we also show robustness to alternative transformations (winsorization at the 99^{th} percentile, adding 0.1, 10, and 100 to earnings prior to logging, and the inverse hyperbolic sine).

We also pre-specified exploratory analyses on: (1) the number of jobs and length of jobs to assess job stability and match quality, and (2) the industry of employment to assess whether letters help youth find jobs in which they now have experience (i.e., those over-represented in SYEP jobs) or whether the letters help market youths' skills to the higher-paying industries that are under-represented in SYEP jobs (see a discussion of these industry definitions in Gelber, Isen, and Kessler (2016)). For (1), we define a job spell as all consecutive quarters worked at the same employer. Other outcomes related to spell length and industry are discussed in Appendix A.1.2 and A.1.3.

3.1.3 Job Application Sample

We randomly selected 2,000 control youth and 2,000 treatment youth from our main 2017 cohort to invite to apply to our job application.[18] Table A.5 shows baseline balance for this subsample. Although the vast majority of baseline covariates are balanced, we note that the treatment group in this subsample is significantly more Hispanic by chance (33 percent in the treatment group versus 29 percent in the control group). As we show in Appendix Section A.6, labor market impacts for Hispanic youth are larger than for other groups. As

youth might have somewhat different effects than those we estimate here.

18. We also invited 1,000 youth from unopened surveys (i.e., outside of our main sample) to ensure that job application behavior was not dramatically different for the youth excluded from our main sample.

a result, the point estimates for employment and earnings are considerably larger for this sample, although the smaller sample size makes the estimates imprecise (see discussion in Appendix Section A.2).

3.1.4 Job Application Outcomes

For the sub-sample of individuals we randomly selected to receive our job application advertisement, we pre-specified three key outcomes: whether someone applied, whether they uploaded a letter, and whether they checked the box to apply to a more selective job as a measure of confidence.

Observing whether there is a treatment-control difference in application rates helps us to test whether there is a supply-side job search response behind any potential changes in labor market outcomes. The proclivity to opt into consideration for the "more selective" job tests for treatment-control differences in self-efficacy and motivation or in beliefs about the probability of being hired. Whether applicants uploaded a letter provides a measure of how much letter use actually changed in job applications.

We also report two additional outcomes to provide a more complete picture of job application behavior: whether someone clicked the link to view the job application (regardless of whether they applied), and whether someone uploaded any file (e.g., CV, transcript, or anything else) in support of their application.

3.1.5 Rated Youth Sample

To distinguish whether the letters provide a general signal or convey useful information about worker productivity, we report labor market impacts separately for youth who receive low overall ratings (categories 1–4: "Very Poor," " Poor," "Neutral," and "Good") versus high overall ratings (categories 5–7: "Very Good," "Excellent," and "Exceptional") from supervisors on the question about overall performance, asked of both treatment and control youth. Unlike our main sample, however, there is the potential for selection into who receives a rating based on supervisor behavior in the survey. Because the survey was designed to maximize the number of letters generated, treatment youth were listed first, along with a longer, multi-page set of questions on each youth; control youth were all listed at the end of the survey on a single page, with check boxes that allowed the supervisor to quickly answer the single overall quality question about each control youth. The different positioning and survey content for treatment and control youth could change the probability a supervisor rated a particular youth. Additionally, supervisors were told (and could decide whether) their responses would be turned into a letter for treatment youth, but not for control youth. The possibility of sending a letter may itself lead supervisors to make different decisions about whether to rate a youth or which rating to give. Because of both differences, we would not necessarily expect the distribution of treatment and control youth to be identical conditional on having a rating or receiving a particular rating.

In fact, treatment youth are significantly less likely to have received a rating than control youth (66 versus 71 percent, p<.01). Although the distribution of baseline characteristics is not statistically different across groups conditional on having a rating, it is significantly different for youth receiving a low rating (Table A.6 shows that treatment youth receiving a low rating are observably different from control youth receiving a low rating, perhaps because supervisors were more hesitant to give low ratings when a letter might be produced than when they knew it would not).

To minimize the role of selection introduced by whether a youth is rated, when we explore treatment effects by ratings, we focus on the sub-sample of youth who were on a survey in which the supervisor rated every treatment youth and every control youth in the survey. There are 13,911 youth who were on such a survey. Since everyone is rated, these surveys leave no room for treatment-control differences in who is rated within the survey. In this group, treatment youth are only 1 percentage point less likely to appear on a completed survey overall (31.6 versus 32.5 percent, p=0.066), with observables jointly balanced (p=0.865). Appendix Table A.8 shows the distribution of baseline characteristics is also similar within rating groups for this sub-sample. Appendix Section A.3 shows that even without this sample restriction, results are very similar when using all youth with a rating.

3.1.6 Education Sample

Because we knew much less about what education data would be available to us at the time of pre-specification, the education analysis is where we deviate most from our pre-analysis plan. As reported above, about 70 percent of our sample has any active record in the DOE records during the period our data cover (2015–2020). But in practice, many of these students either graduated or left school prior to our 2016 and 2017 study years. And while charter school students do appear in DOE data as having active records, DOE does not share any information about school engagement, performance, or graduation for charter school students with outside researchers. Because of the amount of missing data, including on individual elements like GPA and college enrollment even within individuals that have some educational records, we leave the analysis of separate educational outcomes to Appendix Section A.4. That analysis focuses on students who were enrolled in grades 8–12 in the year prior to randomization, were not enrolled in a charter school at the end of the pre-randomization year, and who had not yet graduated. This is the sample who we expect to be in a DOE high school in the year after SYEP (if they do not transfer or stop attending school).

In the main text, we focus on the most substantively important high school outcome, and the one that provides the greatest contribution to the question of how work matters for schooling: high school graduation. We note that because this focus was not pre-specified, it should be considered exploratory. Graduation data are not available for everyone. Per state standards, DOE only reports graduation in the academic years that correspond to a student's on-time (4th), 5th, or 6th year graduation cohort (even if a student returns to school after their 6th year). Graduation data are missing for students who transfer to a charter school; move out of district; fall under another exclusion, such as having an individualized education plan (IEP); or were not in a 4th–6th year graduating cohort between fall 2015 summer 2020. To avoid conditioning the graduation sample on what could potentially be an outcome (e.g., transferring in or out of the District), we restrict the sample to those most likely to be observed in the graduation data based only on pre-randomization characteristics.

In particular, the main text limits attention to students who were enrolled in grades 10–12 in the year prior to randomization, were not enrolled in a charter school at the end of the pre-randomization year, and who had not yet graduated. This is the part of the high school sample with the most available graduation information prior to the end of the data: within this group, 64 percent of students are old enough to have complete graduation data, and all others have reached their 5th-year graduation date by the end of the data.[19] This sample excludes students outside of the DOE, pre-randomization dropouts and graduates, students who temporarily stopped attending public school or had not yet joined the school district in the year before randomization, and those too young to observe their full graduation data (8th and 9th graders in both study cohorts). Appendix Section A.5 reports results including the 8th and 9th graders for whom we can at least observe on-time (but not later) graduation before the end of the data, and Appendix Figure A.8 diagrams the available graduation data by grade and study cohort.

Our main 10th–12th grade graduation sample contains 13,732 students, with no treatment-control difference on being in this sample either overall or conditional on being in our main education data ($\beta = -0.0006, p = 0.926$). We note that there is some chance imbalance on observables within the education data, discussed in more detail in Appendix A.4.3. One benefit of the post-double-selection LASSO is that it adjusts for chance imbalance in a principled way.

19. We assess graduation cohorts based on the grade a student was in during the pre-randomization year, since we only observe the state-defined cohort that officially determines graduation years if someone appears in the graduation data—a potentially selected group. Within our main graduation sample, only 10th graders in the 2017 cohort have incomplete graduation information (n=4,984 have not yet reached their 6th-year graduation date).

3.1.7 Education Outcomes

We explore three main outcomes in our graduation sample. The first measures whether youth graduated on-time (i.e., within four years of starting 9th grade) at any time during our data. The second measures whether someone graduated within six years of starting 9th grade. The third measures "school persistence," which captures whether someone has either graduated or is still attending school in the last academic year of our data (2019–2020)

For our measures, we count any graduation outcome between the start of the academic year when randomization occurred and the end of our graduation data. This will capture most, but not all, eventual high school graduation. There are two reasons why we miss some eventual high school graduation. The first is that, for the 4,984 youth who are in 10th grade prior to the 2017 summer, our data end before their 6th-year graduation date. The second is that graduation after the 6th year is not recorded in our DOE data, so we will not observe any eventual graduation of students who spend more than 6 years in high school. Our school persistence measure is designed to include these youth, who are still pursuing a diploma.[20]

In the main text, we focus on the relationship between educational attainment and labor force involvement. To measure this relationship, we define a set of mutually exclusive joint outcome indicators: working and graduating, never working but graduating, working and not graduating, and never working and not graduating. We define these indicators for all three of our education attainment measures: on-time graduation, ever graduating, and graduating or still attending school. Note that these joint outcomes measure employment over a 2-year follow-up, while graduation includes either a 3- or 4-year follow-up, depending on the study cohort. The pattern across these outcomes will allow us to assess whether any potential shifts in educational attainment occur among the same group that experiences shifts in employment. Results on the three educational attainment indicators separately, as well as other high school performance measures and on-time college enrollment including the full sample of 8th–12th graders, are in Appendix Section A.4.

20. Sixty-four youth in our graduation sample do not have any graduation information available, likely because they transferred out of the district (or joined a different group excluded from state graduation counts) after randomization. Since these individuals did not receive a diploma from NYC DOE, we assign them zeros for graduation. DOE discharge codes suggest there is no treatment effect on whether students transfer out of the district ($\beta = 0.003$, $p = 0.322$, with a control mean of 0.033). Since we do not observe graduation outside the district, the balance on transfers helps to rule out the possibility of differential mobility biasing the graduation results. In theory, differential mobility could also be an issue for our labor market results, since we only observe UI data within New York state. Although the available post-secondary data is limited to a subset of the full sample, the on-time college enrollment measure discussed in the appendix can help assess whether out-of-state mobility is different across treatment groups. That measure records if someone is enrolled in an out-of-state college 6 months after their on-time graduation date, and it suggests that treatment youth are no more likely to leave New York State for college ($\beta = -0.002$, $p = 0.692$, with a control mean of 0.064).

3.2 Analytical Method

3.2.1 Main Analysis

We begin with an intent-to-treat (ITT) analysis by regressing each outcome variable on a treatment indicator and baseline covariates:

$$Y_{it} = \alpha + \beta T_i + \gamma X_{it-1} + \epsilon_{it}$$

where Y_{it} is an outcome for individual i at time t, T_i is an indicator for random assignment to treatment, and X_{it-1} is a vector of covariates measured at or before the time of random assignment. As pre-specified, we use a post-double-selection LASSO to select which covariates to include in each regression (Belloni, Chernozhukov, and Hansen 2014a, 2014b; Belloni et al. 2012).[21] We always include an indicator variable for study cohort, since randomization occurred separately by study year. Because 1,776 individuals appear more than once in the data, we cluster our standard errors on individual as identified by SSN in the SYEP data.

Not every treatment youth on a started survey was sent a letter, either because: they were on a survey answered by someone who was not their direct supervisor, the supervisor did not want to provide a letter, or the supervisor provided ratings that were not positive enough for a letter to be sent. As a result of this kind of non-compliance, the ITT will understate the effect of being sent a letter. In addition, we cannot observe who actually views or uses the recommendation letters in practice. Instead, we do two things to provide a sense of the effect's magnitude for those who are actually treated. First, we use random assignment as an instrument for whether a youth was sent a letter. Since we perfectly observe whether every youth was sent a researcher-generated letter, we can estimate this treatment-on-the-treated effect for everyone. We report control complier means as a baseline measure to assess proportional changes for compliers (Kling, Liebman, and Katz 2007).

Second, we use our job application data to estimate the proportion of treatment youth who actually use the letter in practice. Because we find evidence that letters do not generate a supply-side response, it is possible that the letters work *only* when youth actually show them to employers. In this case, we can approximate the treatment-control difference in letter use with our job application data and use that as an implied first stage to scale the ITT effect. Because our job posting is not representative of all job applications, this extrapolation involves a strong assumption that letter use in the rest of the labor market looks like letter use in our job application. This assumption could fail in two ways: either

21. We implement this with the Stata commands pdslasso and ivlasso (Ahrens, Hansen, and Schaffer 2020). See Appendix Section A.9 for a list of the covariates we offer the LASSO, and for results without any covariates or with all covariates as robustness checks.

because it is easier to remember the letter or submit the letter in our application than in other real-world applications (in which case we would likely understate this LATE effect), or because treatment changes the composition of who applies to our job posting by changing whether youth are employed when we send our advertisement (the direction of which depends on employment treatment effects). We discuss the interpretation issues further below, but in general we consider this a rough approximation of the effect of actually using the letter for those who choose to use it, not a direct estimate of the relevant LATE.

3.2.2 Heterogeneity analysis

Although we pre-specified at the outset that we would not have enough statistical power to differentiate heterogeneous treatment effects, we follow our pre-analysis plan in conducting exploratory analyses based on the characteristics most likely to affect youth's labor market prospects. For all heterogeneity tests, we report the ITT, the first stage, and the IV separately for each group to show how much of ITT differences are from different rates of receiving a letter and how much are from different responses conditional on being sent a letter.

The main text focuses on separating labor market impacts for White and minority youth, where minority is defined as any non-White self-classification, including Black, Hispanic, Asian, and Other (including American Indian, Pacific Islander, mixed race, or unspecified other). To help identify whether employers are responding to the specific information in the letters, we also test for heterogeneity across supervisor ratings. If the letters are successfully communicating specific information, we would expect that providing letters with higher ratings would generate larger labor market benefits.

Appendix Section A.6 breaks down effects by specific racial and ethnic groups, and it reports heterogeneity on the other pre-specified categories: age, gender, school enrollment (as self-reported on the application to SYEP), and neighborhood. The appendix also explores heterogeneity by previous work experience to see if the effects are limited to those who lack other signals of an ability to get a job on their applications.

4 Results

4.1 Summary Statistics

Table 1 shows average pre-randomization characteristics for the treatment and control groups. No more differences are significantly different than would be expected by chance, nor are they jointly significantly different ($p = 0.267$). Study participants reflect the population that participates in NYC's SYEP. On average, they are just over 17 years old, about 43 percent male, largely identify as minorities (only 12.5 percent list being White on their application), and 75 percent report being in high school in the spring prior to the SYEP. About 45 percent of participants never appear in the UI data prior to their participation in SYEP, but 97 percent work during the SYEP year, earning an average of just under $2,400 that year.

4.2 Labor Market Effects

Table 2 reports the main labor market effects. Panel A shows that being assigned to the treatment group increases employment rates by 1.3 percentage points (1.8 percent relative to the control mean of 70 percent) during the year following letter distribution.[22] Actually being sent the letter increases year 1 employment by 3.1 percentage points (4.5 percent relative to the control complier mean). The point estimates in the second year after letter distribution are still positive but smaller and not statistically significant. However, the increase in employment is still significant over the cumulative two-year follow-up: being sent a letter increases net two-year employment rates by 2 percentage points (2.3 percent).

It is likely that the employment effect will fade out eventually, since almost all control youth will eventually work in the formal labor market at least once. But earnings changes would not necessarily fade out if the letter is helping set youth on a better employment trajectory or find better jobs. Panels B and C report program effects on winsorized earnings in dollars and log(earnings+1), respectively. The treatment effect grows in levels over the two years observed, though effects are somewhat noisy. Those sent a letter earned a total of $433 more over the two-year follow-up period ($p = 0.101$), a 4.4 percent increase. The $log+1$ transformation is more precise, with Panel C showing a significant 18.6 percent increase in cumulative earnings and significant increases in both years 1 and 2.

Because there is a treatment effect on the extensive margin, the results may be sensitive to how we handle the proportional change at $0. Table A.1 shows alternative level, log, and asinh transformations. The results suggest that the magnitude of the change is somewhat sensitive to functional form, ranging up to 23 percent, but generally statistically significant. We focus on the 4.4 percent estimate in the main text both because our pre-specified primary outcome was winsorized earnings and because it is the most conservative estimate. Since we also pre-specified that we would use a range of robustness checks to adjust for skewness, we conclude that the evidence suggests that the letters of recommendation generated a sizable increase in earnings.

Table 3 digs more deeply into the UI records to understand how labor market outcomes are improving. The first column shows that treatment youth work in 0.05 more quarters (0.11 for letter recipients) than their control counterparts. The last column shows that conditional

22. Letters were sent in December of 2016 and November of 2017, so we include the final quarter of each calendar year as the first quarter of year 1.

on working at all (i.e., for those selected into work), treatment youth find jobs sooner than control youth (0.12 quarters sooner for letter recipients). Together, this pattern suggests that the letters help youth shorten the job search process, but do not merely substitute early work for later work; youth work more than they would have otherwise.

One concern about this pattern is the possibility that supervisors could be over-updating their beliefs about youth, interpreting the letters as a stronger quality signal than they actually are. If so, we might expect increased churn, with treatment youth getting hired and fired more frequently than controls. The rest of Table 3 suggests this is not the case: there is no increase in the number of job spells treatment youth have. The point estimate on the number of jobs (including 0s) is positive but not statistically significant, partly capturing the change at the extensive margin. Conditional on working at all (column 3), which is a selected sample, the point estimate is a precise zero. In other words, there is no evidence of additional churn among those who work. And as we would expect for young people who start working earlier and work more in the same number of jobs, Appendix Section A.1.2 shows that some job spells get longer. This provides further evidence that letters are not reducing—and in fact may be increasing—the quality of job matches.

4.3 Mechanisms

4.3.1 Assessing Changes in Labor Supply

A key question about the observed increase in labor market success among treatment youth is whether the letters increase labor supply by increasing youth job search intensity or confidence, or whether the letters increase labor demand by changing beliefs about applicants with letters or increasing the salience of those applicants among employers. By distributing our own job posting to 4,000 treatment and control youth, we are able to generate some evidence on why the letters increase employment and earnings and to approximate how treatment changes letter use in job applications. Appendix Section A.2 shows that we have baseline balance within this sample and shows the main employment results for this group.

Table 4 suggests that supply-side responses (increased job search, motivation, or confidence) are unlikely to be driving the labor market improvements. We find no evidence that treatment youth are more likely to click on the application link or actually apply to our posting.[23] The second column shows that 8.8 percent of the control group and 8.2 percent of the treatment group applied to our job, a difference that is not statistically significant.

23. The "applied" variable here measures whether a youth entered enough information in the application for us to know who filled out the application form. We define "applied" this way because we hired people even if they did not answer all the questions on the application. To actually be hired, the youth additionally needed to click submit on the final page of the application. There is also no treatment-control difference on whether youth were hired per this definition.

We also find no evidence that the letter increased confidence among applicants conditional on applying; treatment youth are no more likely to volunteer for the more selective job than control youth (see the last column of Table 4, which, adjusting for application rates, translates into 60 percent of control applicants and 51 percent of treatment applicants checking the box to apply for the more selective job).

Of course, it is possible that even though the letters did not change the rate at which young people applied to our job, they could have changed the composition of who applied. Since treatment youth were more likely to be employed in the formal labor market, their interest in our short-term, online job may have been directly affected by treatment (even though we framed the job as flexible enough to be compatible with other work). That said, we cannot reject the null that observables are jointly unrelated to treatment status among applicants, suggesting this is not likely to be the case.[24] Additionally, even if we condition on not being employed elsewhere during the quarter the job application was distributed (a selected group), there is still no significant difference in application rates or our confidence measure ($\beta = -0.01$, $p = 0.351$ for applying and $\beta = -0.01$, $p = 0.132$ for checking the selective box).[25] So the lack of an increase in supply-side job-seeking behavior does not appear to be due to treatment youth being more likely to be employed already. Overall, the evidence from our job application suggests that labor market improvements are coming from employers responding to letters of recommendation, not from changes in youth's application behavior or confidence.

4.3.2 Assessing the First Stage

As an important check on whether treatment youth actually use the letters we send them—a necessary condition for employers to be able to respond to the letters—the final two columns of Table 4 show treatment effects on the files job applicants uploaded in their application to our job posting. There is no detectable change by treatment in the probability that youth upload some form of supporting material. But there is a dramatic change in whether youth upload a letter of recommendation. Only 0.4 percent of the control group submits a letter, including zeros for those who do not apply (conditional on applying, this translates to 4.5 percent of control applicants submitting a non-intervention letter with their application). Treatment youth are two and a half times more likely to submit a letter of recommendation than the control group: 1.4 percent of all those invited to apply submit a letter (16.5 percent

24. We test for differences between the treatment and control individuals who applied for our job, conditional on being in our application sample, by interacting each baseline covariate with an indicator for whether the individual applied, regressing treatment on all covariates and these interactions, and then testing the hypothesis that all interaction coefficients are jointly 0. The p-value of this test is 0.14.

25. The same is true conditional on being employed in that quarter: $\beta = 0.0008$, $p = 0.959$ for applying.

conditional on applying). Since about 40 percent of treatment youth actually received a letter, this implies that about 41 percent of letter recipients use them when they apply to a job (16.5 percent relative to 40 percent).

Given the lack of supply side response, it seems reasonable to suppose that letters might only work when employers see them. If so, the observed rates at which letters are used can also benchmark the first stage of letter use, which under the exclusion restriction we can use to extrapolate how big employment responses are for youth who actually use their letters. If we make the quite strong assumptions that the difference in letter use we observe in our job application applies to the entire sample, that treatment and control youth apply to jobs at the same rate, and that everyone applies to at least one job, then the implied first stage for letter use is a 12 percentage point increase (4.5 versus 16.5 percent among applicants). Scaling our main ITT effects by this first stage would in turn imply that the employment increase for those who use the letter is about 15 percent relative to baseline in the first year, and 8 percent over two years, with an additional $1,400 in earnings over that time.

Because of all the extrapolation involved in this calculation, we view it as a back-of-the-envelope estimate. If we think that it might be easier to use the letter in our application than in typical job applications (e.g., because receiving a job advertisement that references SYEP and having a screen to upload supporting material reminds treatment youth about the letter or primes them to use it more than in a typical job application), or that not everyone applies to at least one job, then we are likely overstating the number of treatment youth who used a letter relative to controls. In that case, our 12 percentage point first stage would be an overestimate, and the actual LATE would be even larger than our calculations here suggest.

4.3.3 Assessing Changes in Labor Demand

The evidence so far suggests that employers are the ones responding to the increased use of letters of recommendation in the job applications of treated youth. A final mechanism question is how those letters affect their hiring behavior. It is possible that because letters are infrequently included in typical job applications, it is the presence of the letter itself—regardless of content—that makes an application more salient, resolves some basic uncertainty about whether an applicant is likely to show up at all, or overcomes statistical discrimination. Alternatively, employers may be using the content of the letter to try to discern something more nuanced about future employee reliability or productivity.

Although we did not send letters where SYEP supervisors included too few positive comments about the youth they supervised, there is still variation in how positive supervisors were in their letters that allows us to assess whether employers respond to letter content. Table 5 shows employment and earnings effects separately for youth who received low ratings (categories 1–4, corresponding to "Very Poor," " Poor," "Neutral," and "Good") and high ratings (categories 5–7, corresponding to "Very Good," "Excellent," and "Exceptional").[26] Highly rated youth were much more likely to receive a letter (81 percent versus 33 percent). So the ITT differences between the groups reflect both differences in letter receipt and differences in outcomes conditional on being sent a letter, although the substantive pattern of results is relatively similar for both the ITT and TOT.

The first result of note is that the ratings do seem to capture attributes that matter in the labor market. Looking at the control means, low-rated youth are 6 percent less likely to work during year 1 (67 versus 72 percent employed), though they catch up to high-rated youth over time. They also earn just under $1,500 (14.6 percent) less over 2 years. Second, we find that, cumulatively, the low-rated group has net employment effects close to 0 and cumulative earnings point estimates that are negative but with huge standard errors. In contrast, the high-rated group has employment and earnings effects that grow over time, such that they are driving basically all of the net positive effects of the treatment.

It appears, then, that employers are using the substance of the letters to identify those who are likely to be highly productive employees, but who might not otherwise be noticed during the hiring process. While one might wonder whether the low-rated group simply chooses not to use letters in their job applications, results from our job application suggest otherwise (see Appendix Section A.3). For every 100 letters sent to high-rated treatment youth, we received 3 job applications that included letters. For every 100 letters sent to low-rated treatment youth, we received 4 applications including letters. So there is no indication that low-rated letter recipients are less likely to use letters when applying for jobs.[27] The group of young people who did not impress their SYEP supervisors as much may need more intensive investment in improving skills to reap long-term gains.

4.4 Work and Graduation

Education results on engagement and performance outcomes are in Appendix Section A.4. In general, there is little evidence that letters improve student performance in school (e.g., by changing teacher or guidance counselor beliefs or encouraging college application). While none of the treatment effects are statistically significant for the full education sample, there is one pattern that becomes significant in several subgroups and alternative specifications: on-time (4-year) graduation shows a substantively important decline, while the point estimates

26. Note that if youth received an overall rating less than "Good," the paragraph that included text about the overall rating was not printed in the letter. These letters were still produced, though, as long as enough other attributes were rated positively enough.

27. While high-rated letter recipients apply at somewhat higher rates and use letters somewhat more often, many more of them are sent letters than their low-rated counterparts.

on ever graduating (including delayed graduation), school persistence (graduating or still attending), and enrolling in college immediately after 4-year graduation, are much closer to zero. This pattern provides suggestive evidence that, on average, recommendation letters increase employment while slowing down—but not stopping—high school graduation for those still in school.

A natural hypothesis is that our employment and education effects are driven by the same group of youth: that by pulling young people into the labor force, the letters make it harder for them to complete their high school education on time, leading them to spend additional time in school. Table 6 tests this hypothesis by showing IV effects for the joint outcomes of working and graduating, never working but graduating, working but not graduating, and never working nor graduating. Appendix Table A.13 is the ITT version of this table. Panel A shows these outcomes using on-time graduation; Panel B uses whether someone ever graduated—on time or otherwise—within our data; and Panel C shows school persistence (i.e., whether someone graduated or is still attending school in the final year of our data).

The pattern of results is consistent with letters reducing on-time graduation—but not eventual graduation—for the same group of youth who also have positive employment impacts from the letters. The second column of Panel A shows that treatment significantly decreases the proportion of youth who graduate on time without working by 2.1 percentage points for letter recipients. Treatment also generates a corresponding increase, of roughly similar size (3.0 percentage points), in the proportion of youth who work but do not graduate on time, as shown in column 3. Since there are no significant changes in the other two categories, this is suggestive of a shift from graduating on time without work to working but failing to graduate on time. The additional 0.9 percentage point increase in working without graduating on time appears to come from the group that neither works nor graduates on time, as shown in column 4, suggesting a modest employment increase from youth who would not have graduated on time even if they had not received the letter.

By contrast, Panel B shows that for ever graduating, there is no significant change in the proportion of youth who work but do not graduate. Rather, treatment generates a significant 2.6 percentage point decrease in the proportion of letter recipients who never work but graduate, and an offsetting increase of 2.4 percentage points in the proportion who both work and graduate. In other words, the letters seem to encourage employment among those who graduate. The point estimates among non-graduates show a similar pattern of a negative point estimate for not working with an offsetting positive point estimate for working. These estimates, however, are substantially smaller and not statistically significant.

Panel C measures educational attainment as school persistence: either graduating or continuing to attend school. The basic pattern of results is similar to the results in Panel B.

Among those who do not persist, we see a decline in never working and an offsetting increase in working. We see the same pattern among those who do persist.

Appendix Table A.16 shows that these shifts are concentrated among students who have below-median GPAs in the year before randomization, consistent with the idea that it is students struggling in school who are most responsive to the letter and whose educational attainment is most harmed by the increase in work. Appendix Section A.5 discusses these and additional robustness checks, including showing a similar but slightly noisier set of results when 8th and 9th graders, who have had less time to graduate, are included in the analysis.

In sum, for the subset of study youth for whom we have the most complete graduation data, this joint outcome analysis suggests that the shock to employment generated by the letters slows down graduation, but does not stop it. These results provide a useful addition to the literature on working during school, which typically has been unable to measure on-time graduation and has hit a ceiling effect when analyzing overall graduation (Buscha et al. 2012; Staff, Schulenberg, and Bachman 2010; Monahan, Lee, and Steinberg 2011; Ruhm 1997; Baum and Ruhm 2016). The welfare implications of slowing down graduation—and whether the slightly longer high school career outweighs the benefits of the additional work experience and earnings—depends on how long and by how much the letters affect the trajectory of future longer-term outcomes.

It is worth emphasizing that only about 30 percent of our overall sample is in this graduation analysis, due to the smaller set of youth who are of relevant age and for whom we have education data. The rest of the youth in our study are either too young for us to observe graduation, are out of school already, or are enrolled in schools that are not in our data. Those not in our high school data still have significant increases in year 1 employment as well as much larger point estimates on earnings than the high school sample (see Appendix Section A.4). So, from a policy perspective, it may be feasible to focus on mitigating information frictions that impact youth who are not still in high school, or at least students not on the margin of graduating on-time.

4.5 Heterogeneity

Table 7 shows that, while we do not have the statistical power to differentiate between the two groups, the labor market effects of the letters appear to be driven by racial and ethnic minorities. The letters have no significant effect on White youth, who show negative but imprecise point estimates from the letters. Effects are only positive and statistically different from zero for the non-White (Black, Hispanic, Asian, and Other) youth in the sample.[28]

28. Appendix Section A.6 shows results separately for the individual racial and ethnic groups, as well as for other subgroups of interest.

The first-stage results suggest part of the difference in the ITT effects is that minority youth are much more likely to be sent a letter (42 percent versus 30 percent for minority youth and White youth respectively). This difference may have to do with differential selection into the SYEP in the first place, since only 12.5 percent of our SYEP participants are White, or with differences in the kinds of SYEP supervisors for whom minority and White youth work. But the IV results show that, even conditional on being sent a letter, the point estimates for employment and earnings are much bigger for minority youth than for White youth. One might wonder whether the larger IV effect reflects differences in letter quality; perhaps letters for minority youth matter more because they are stronger letters. However, we observe that the opposite is true: conditional on receiving a letter, White youth receive ratings that are 0.43 points higher (on our 7-point scale) than minority youth, with no significant differences in whether they use the letter on our job application (see Appendix Section A.7 for descriptions of how letters and job application behavior differ by subgroup). Consequently, it seems that the letters sent to minority youth have a particularly powerful effect because of how employers respond to them.

Table 8 shows that the education results in Table 6 are, in fact, driven by minority students. They are driving the declines in on-time graduation, although the proportion who both work and graduate eventually still increases. The shifts in persistence are consistent with the possibility that the letters help both those who would and would not finish school shift into the labor market. But since multiple outcomes move at once, we cannot rule out the possibility that some youth may be pulled out of school by their increased employment. Interestingly, the substantive patterns are largely in the opposite direction among White students, with hints that the letters are helping White students' educational success without increasing employment. But the small number of White youth in the data means the changes are not statistically distinguishable from zero.

5 Discussion

Sending youth a few copies of a letter and an email with a link to that letter increased employment rates by 4 percent—and perhaps as much as 15 percent for those who actually use the letter. These results provide new evidence that there are, in fact, frictions in the labor market for youth, and minority youth in particular, that are relatively low cost to overcome. We do not find differences in job-seeking behavior among treatment youth other than using the letter, suggesting that employers are the ones responding to the additional information contained in the letters. This interpretation is bolstered by results showing that higher performing youth get a larger labor market benefit from the letters.

We also find that recommendation letters lead to a decline in on-time graduation, driven by substitution toward work on the margin. In addition to being important for any future policy decisions about letters of recommendation or other signals about youth, this finding also speaks to the literature on the impact of working during high school. Our letters provide a plausibly random shock to working during high school, which appears to extend the time spent in high school. To assess the welfare implications of this letter-generated substitution, we would need to make some strong assumptions about how long the increase in earnings will last and how that compares to the longer time in high school. It seems likely that the net effect may not be beneficial, especially if any subgroup leaves school entirely. Reducing employment frictions is most likely to have a net benefit for those who have already finished their high school careers, or at least are not currently on the margin of graduating (although we find few labor market benefits among high-achieving students).

Overall, the labor market results indicate that employers respond to credible information about youth, such that finding additional ways to provide them with personalized information about an individuals' strengths could help improve labor market outcomes among low-income, largely minority populations like those in our study. For social programs looking to help youth or other disadvantaged groups capitalize on their training or early work experience, this is an important insight.

A natural question is how broadly this finding applies—whether we are documenting that letters help overcome the particular stigma associated with SYEP participation or that, more broadly, at least some youth unemployment is due to frictions surrounding the availability of information about young applicants. The answer to this question rests partly on whether the employers in our data knew that youth applicants were SYEP participants, which would be necessary for the stigma story. While we cannot observe that directly, we can take a hint from the applications that youth submitted in response to our job advertisement. In those applications, only 22 percent of applicants self-identify as a SYEP participant in either their list of work experience or their résumé. Given that almost 80 percent of job applicants would not appear to employers as SYEP participants—and that the recommendation letters came on letterhead from the agency that runs the program, increasing the salience of the SYEP—it seems plausible that the frictions we document are not specific to SYEP-related beliefs among employers.

Despite the positive effects of the letters on labor market outcomes, our findings do not necessarily imply that policymakers should try to give everyone letters of recommendation. Our estimates are for youth who receive letters when survey responses are voluntary and responses are positive enough. Effects may differ outside of this population.[29] In addition,

29. It is difficult to say from the observable differences in youth across the opened and unopened surveys whether effects would be bigger or smaller if supervisors were forced to fill out the surveys. The unopened

any effort to generate more widespread use of credible signals like letters of recommendation could result in displacement; youth with letters might gain jobs, but at the expense of those who would otherwise have taken those jobs.

Such displacement and general equilibrium effects are worth considering as part of efforts to scale up such programs. There are several conditions under which a scaled-up version could be beneficial, even with considerable displacement. If policymakers valued equity, then transferring job opportunities to those farther down the income distribution or to historically marginalized groups might be socially beneficial. Alternatively, even with no net change in employment, letters could generate efficiency gains by helping employers and employees find better matches. And, if employers end up leaving some vacancies open in the face of too much uncertainty about applicants, as they appear to in an online marketplace (Pallais 2014), a widespread information-sharing intervention might increase overall employment. Finally, there could also be general equilibrium effects on the supply side; if young people understand that they may receive helpful recommendation letters, they may work harder in their jobs, generating additional productivity as well as better letters to which future employers will respond more positively.

Even in partial equilibrium, our experiment establishes that information frictions prevent minority young people from getting jobs they could otherwise succeed in. Further research into the precise way employers update their beliefs or substitute across workers in response to efforts to mitigate these frictions would be a productive next step in assessing the most effective way to leverage our findings into higher youth employment.

surveys contained more White youth, who have smaller labor market effects. But they also had more youth already out of high school, which could diminish graduation crowd-out, and more youth with work experience prior to SYEP, who have directionally larger point estimates, see Appendix Section A.6.

References

Abel, Martin, Rulof Burger, and Patrizio Piraino. 2020. "The Value of Reference Letters: Experimental Evidence from South Africa." *American Economic Journal: Applied Economics* 12, no. 3 (July): 40–71. Accessed April 5, 2021.

Agan, Amanda, and Sonja Starr. 2018. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment*." *The Quarterly Journal of Economics* 133, no. 1 (February 1, 2018): 191–235. Accessed February 22, 2019.

Ahrens, Achim, Christian Hansen, and Mark Schaffer. 2020. *pdslasso and ivlasso: Progams for post-selection and post-regularization OLS or IV estimation and inference.* http://ideas.repec.org/c/boc/bocode/s458459.html.

Bartik, Alexander, and Scott Nelson. 2019. "Deleting a Signal: Evidence from Pre-Employment Credit Checks." *University of Chicago, Becker Friedman Institute for Economics Working Paper,* nos. 2019-137.

Baum, Charles L., and Christopher J. Ruhm. 2016. "The Changing Benefits of Early Work Experience." *Southern Economic Journal* 83, no. 2 (October): 343–363.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen. 2012. "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain." *Econometrica* 80 (6): 2369–2429. ISSN: 1468-0262, accessed October 29, 2021.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014a. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28, no. 2 (May): 29–50. ISSN: 0895-3309, accessed October 29, 2021.

———. 2014b. "Inference on Treatment Effects after Selection among High-Dimensional Controls†." *The Review of Economic Studies* 81, no. 2 (April 1, 2014): 608–650. ISSN: 0034-6527, accessed October 29, 2021.

Bertrand, Marianne, and Esther Duflo. 2017. "Field experiments on discrimination." *Handbook of economic field experiments* 1:309–393.

Bureau of Labor Statistics, U.S. Department of Labor. 2020. *Employment status of the civilian noninstitutional population by age, sex, and race.* Accessed April 5, 2021.

———. 2021. *Employment and Unemployment Among Youth — Summer 2021.* Accessed November 29, 2021.

Buscha, Franz, Arnaud Maurel, Lionel Page, and Stefan Speckesser. 2012. "The Effect of Employment while in High School on Educational Attainment: A Conditional Difference-in-Differences Approach*." *Oxford Bulletin of Economics and Statistics* 74 (3): 380–396.

Card, David, Jochen Kluve, and Andrea Weber. 2018. "What works? A meta analysis of recent active labor market program evaluations." *Journal of the European Economic Association* 16 (3): 894–931.

Crépon, Bruno, and Gerard J Van Den Berg. 2016. "Active labor market policies." *Annual Review of Economics* 8:521–546.

Davis, Jonathan M.V., and Sara B. Heller. 2020. "Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs." *The Review of Economics and Statistics* 102, no. 4 (October): 664–677.

DeSilver, Drew. 2021. "In the U.S., teen summer jobs aren't what they used to be." Pew Research Center. Accessed April 5, 2021.

Doleac, Jennifer L., and Benjamin Hansen. 2020. "The Unintended Consequences of "Ban the Box": Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden." Publisher: The University of Chicago Press, *Journal of Labor Economics* 38, no. 2 (April 1, 2020): 321–374. Accessed September 13, 2021.

Gelber, Alexander, Adam Isen, and Judd B. Kessler. 2016. " The Effects of Youth Employment: Evidence from New York City Lotteries *." *The Quarterly Journal of Economics* 131, no. 1 (February): 423–460.

Inanc, Hande. 2020. *Breaking Down the Numbers: What Does COVID-19 Mean for Youth Unemployment?* Mathematica Policy Research Reports. Mathematica Policy Research. Accessed April 5, 2021.

Kaas, Leo, and Christian Manger. 2012. "Ethnic discrimination in Germany's labour market: A field experiment." *German economic review* 13 (1): 1–20.

Kahn, Lisa B. 2010. "The long-term labor market consequences of graduating from college in a bad economy." *Labour Economics* 17, no. 2 (April): 303–316. Accessed February 22, 2019.

Katz, Lawrence F, Jonathan Roth, Richard Hendra, and Kelsey Schaberg. 2020. *Why Do Sectoral Employment Programs Work? Lessons from WorkAdvance.* Technical report. National Bureau of Economic Research.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. “Experimental Analysis of Neighborhood Effects.” *Econometrica* 75, no. 1 (January): 83–119. Accessed February 22, 2019.

Kochhar, Rakesh, and Amanda Barroso. 2020. “Young workers likely to be hard hit as COVID-19 strikes a blow to restaurants and other service sector jobs.” Pew Research Center. Accessed April 5, 2021.

Le Barbanchon, Thomas, Diego Ubfal, and Federico Araya. 2020. “The Effects of Working while in School: Evidence from Uruguayan Lotteries.” *Available at SSRN 3398385.*

Modestino, Alicia Sasser. 2019. “How Do Summer Youth Employment Programs Improve Criminal Justice Outcomes, and for Whom?” *Journal of Policy Analysis and Management* 38, no. 3 (June): 600–628. Accessed September 13, 2021.

Monahan, Kathryn C., Joanna M. Lee, and Laurence Steinberg. 2011. “Revisiting the Impact of Part-Time Work on Adolescent Adjustment: Distinguishing Between Selection and Socialization Using Propensity Score Matching.” *Child Development* 82 (1): 96–112.

Neumark, David. 2002. “Youth Labor Markets in the United States: Shopping Around vs. Staying Put.” *Review of Economics and Statistics* 84, no. 3 (August): 462–482. Accessed February 22, 2019.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz. 2012. “The Short- and Long-Term Career Effects of Graduating in a Recession.” *American Economic Journal: Applied Economics* 4, no. 1 (January): 1–29. Accessed February 22, 2019.

Pallais, Amanda. 2014. “Inefficient Hiring in Entry-Level Labor Markets.” *American Economic Review* 104, no. 11 (November): 3565–3599. Accessed February 22, 2019.

Rosenthal, Robert, and Lenore Jacobson. 1968. “Pygmalion in the classroom.” *The Urban Review* 3, no. 1 (September): 16–20. Accessed February 22, 2019.

Ruhm, Christopher J. 1997. “Is High School Employment Consumption or Investment?” *Journal of Labor Economics* 15 (4): 735–776.

Spievack, Natalie, and Nathan Sick. 2019. “The Youth Workforce: A Detailed Picture.” Urban Institute, July 22, 2019. Accessed April 5, 2021.

Staff, Jeremy, John E. Schulenberg, and Jerald G. Bachman. 2010. “Adolescent work intensity, school performance, and academic engagement.” *Sociology of Education* 83 (3): 183–200.

Sum, Andrew, Ishwar Khatiwada, Walter McHugh, and Will Kent. 2014. "Deteriorating Labor Market Fortunes for Young Adults." *Challenge* 57, no. 3 (May 1, 2014): 60–83. Accessed April 5, 2021.

SYEP Annual Summary. 2017. Technical report. New York City Department of Youth and Community Development.

Tables and Figures

Table 1: Descriptive Statistics

	Control Mean	Treatment Mean	Test of Difference
N	21,695	21,714	
Age	17.2	17.2	0.641
Male	0.427	0.427	0.991
Black	0.409	0.411	0.805
Hispanic	0.289	0.289	0.944
Asian	0.129	0.130	0.734
White	0.124	0.125	0.756
Other Race	0.049	0.045	0.080
In High School	0.755	0.751	0.339
HS Graduate	0.044	0.042	0.202
In College	0.173	0.180	0.081
Not in UI Data	0.009	0.011	0.128
Never Employed Pre-SYEP	0.450	0.457	0.113
Ever Worked, Year -4	0.153	0.149	0.210
Earnings, Year -4	318	320	0.882
Ever Worked, Year -3	0.267	0.266	0.840
Earnings, Year -3	585	585	1.000
Ever Worked, Year -2	0.437	0.435	0.627
Earnings, Year -2	1072	1050	0.412
Ever Worked, Year -1	0.966	0.966	0.798
Earnings, Year -1	2379	2368	0.722
No Education Match	0.126	0.123	0.359
In HS Sample	0.454	0.454	0.938
Joint F-Test	F(24, 41632) = 1.16, p=.267		

Notes: N = 43,409. 390 youth missing race/ethnicity and 1 missing education. Test of Difference reports the p-value from a regression of each characteristic on a treatment indicator, controlling for a cohort indicator and using standard errors clustered on individual.

Table 2: Labor Market Effects

Year	1	2	Cumulative
Panel A: Employment			
ITT	0.0127***	0.0058	0.0079**
	(0.0041)	(0.0041)	(0.0034)
CM	0.701	0.72	0.841
Sent Letter (IV)	0.0313***	0.0144	0.0195**
	(0.0102)	(0.0102)	(0.0083)
CCM	0.697	0.728	0.841
Panel B: Earnings, Winsorized at 99.5th Percentile			
ITT	60.03	110.1	168.66
	(45.89)	(73.23)	(106.86)
CM	3579	5964	9543
Sent Letter (IV)	154.11	281.4	433.17
	(113.40)	(180.95)	(264.02)
CCM	3729	6162	9894
Panel C: Log(Earnings + 1)			
ITT	0.095***	0.059*	0.075**
	(0.033)	(0.035)	(0.030)
CM	5.61	6.08	7.33
Sent Letter (IV)	0.234***	0.146*	0.186**
	(0.081)	(0.087)	(0.073)
CCM	5.64	6.18	7.39

Notes: N = 43,409. Winsorization in Panel B recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 3: Amount and Timing of Work

	Num Quarters Worked	Num of Job Spells	Num of Job Spells if >0	Time to First Qtr Worked
ITT	0.045**	0.019	0.002	-0.048**
	(0.022)	(0.014)	(0.014)	(0.020)
CM	3.46	1.98	2.36	2.19
Sent Letter (IV)	0.111**	0.046	0.006	-0.119**
	(0.054)	(0.034)	(0.034)	(0.048)
CCM	3.59	1.98	2.35	2.18
N	43409	43409	36647	36647

Notes: Spells are defined as consecutive quarters with earnings from same employer. Time to First Qtr Worked conditions on having at least one spell. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 4: Job Application Effects

	Clicked Link	Applied	Checked Selective Job Box	Uploaded Any File	Included Letter of Rec
ITT	-0.007	-0.006	-0.01	0.003	0.010***
	(0.009)	(0.009)	(0.007)	(0.007)	(0.003)
CM	0.103	0.088	0.053	0.052	0.004
Sent Letter (IV)	-0.02	-0.019	-0.027	0.006	0.024***
	(0.024)	(0.022)	(0.017)	(0.018)	(0.007)
CCM	0.138	0.123	0.082	0.065	0.009

Notes: N = 4,000. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 5: Labor Market Effects for Youth with High and Low Supervisor Ratings

	First Stage	Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Low Ratings		0.0250*	-0.0146	0.002	63.24	-163.28	-98.75
		(0.0133)	(0.0134)	(0.0110)	(129.42)	(211.89)	(308.35)
ITT, High Ratings		0.013	0.0238***	0.0174**	106.42	338.72**	437.84*
		(0.0087)	(0.0087)	(0.0070)	(99.38)	(164.96)	(237.14)
P-value, test of diff.		0.455	0.016	0.235	0.791	0.062	0.168
CM, Low		0.673	0.721	0.836	3109	5409	8518
CM, High		0.715	0.720	0.846	3729	6251	9979
IV, Low Ratings	0.3301***	0.0756*	-0.0442	0.0056	190.62	-506.09	-311.39
	(0.0103)	(0.0405)	(0.0405)	(0.0332)	(391.67)	(641.82)	(933.60)
IV, High Ratings	0.8108***	0.0161	0.0293***	0.0212**	130.92	413.65**	535.92*
	(0.0057)	(0.0108)	(0.0108)	(0.0087)	(122.57)	(203.44)	(292.50)
P-value, test of diff.	0.000	0.155	0.08	0.649	0.884	0.172	0.387
CCM, Low		0.61	0.756	0.821	3012	5942	8950
CCM, High		0.713	0.717	0.843	3626	6079	9714

Notes: To avoid selection into who is rated within a survey, sample includes only youth who were on a survey where the supervisor rated all listed youth (N = 13,911). Low Ratings includes rating categories 1–4; High Ratings includes rating categories 5–7. P-value, test of diff shows p-values for tests of the null hypothesis that treatment effects are equal in low and high ratings groups. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 6: IV Effects on Joint Employment and Graduation Outcomes

	Panel A: On-Time Graduation			
	Ever Work, On-time Grad	Never Work, On-time Grad	Ever Work, Not On-time	Never Work, Not On-time
Sent Letter (IV)	0.0031	-0.0209*	0.0296***	-0.0105
	(0.0146)	(0.0117)	(0.0103)	(0.0071)
CCM	0.727	0.12	0.114	0.038
	Panel B: Any Graduation			
	Ever Work, Graduated	Never Work, Graduated	Ever Work, Not Graduated	Never Work, Not Graduated
Sent Letter (IV)	0.0238	-0.0257**	0.0065	-0.0048
	(0.0145)	(0.0120)	(0.0092)	(0.0065)
CCM	0.758	0.127	0.086	0.029
	Panel C: Any Graduation or Continued Attendance			
	Ever Work, Persisted	Never Work, Persisted	Ever Work, Not Persisted	Never Work, Not Persisted
Sent Letter (IV)	0.0154	-0.0198	0.0153*	-0.0108*
	(0.0146)	(0.0123)	(0.0086)	(0.0059)
CCM	0.793	0.129	0.051	0.027

Notes: N=13,732. Analysis is conducted on the main graduation sample (non-charter 10th–12th graders in the pre-randomization year, see text for details). First stage for this subsample is 0.44. Panel A shows whether someone ever worked during the two-year follow up and whether they graduated on-time (i.e., 4th-year graduation). Panel B shows whether someone ever worked during the two-year follow up and whether they ever graduated (i.e., 4th-, 5th-, or 6th-year graduation). Panel C shows whether someone ever worked during the two-year follow up and whether they either graduated or had positive days attended in the last year of our data. CCM shows control complier means, which may not total to 1 across categories due to estimation error in the IV and the inclusion of different sets of covariates in the post double-selection LASSO. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 7: Labor Market Effects for Minority and White Youth

	First Stage	Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Minority		0.0134***	0.0066	0.0090**	79.03	149.27*	227.66**
		(0.0044)	(0.0044)	(0.0036)	(48.13)	(77.52)	(112.60)
ITT, White		0.0048	-0.0019	-0.0031	-70.27	-162.15	-230.22
		(0.0114)	(0.0122)	(0.0096)	(144.77)	(218.80)	(328.57)
P-value, test of diff.		0.483	0.513	0.236	0.328	0.18	0.187
CM, Minority		0.6932	0.7229	0.839	3540	5958	9498
CM, White		0.7518	0.6949	0.851	3754	5702	9457
IV, Minority	0.4188***	0.0319***	0.0158	0.0214**	194.27*	365.42**	557.35**
	(0.0036)	(0.0106)	(0.0105)	(0.0086)	(114.85)	(184.98)	(268.66)
IV, White	0.2973***	0.0157	-0.0077	-0.0112	-241.61	-563.96	-798.88
	(0.0088)	(0.0385)	(0.0412)	(0.0323)	(488.27)	(737.83)	(1107.98)
P-value, test of diff.	0.000	0.685	0.58	0.329	0.385	0.222	0.234
CCM, Minority		0.692	0.729	0.839	3644	6082	9728
CCM, White		0.753	0.715	0.865	4406	6611	11011

Notes: N = 37,653 Minority youth and N = 5,366 White youth. 390 observations are dropped due to missing race/ethnicity. Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 8: IV Effects on Joint Employment and Graduation Outcomes for Minority and White Youth

		Panel A: On-Time Graduation			
		Ever Work, On-time	Never Work, On-time	Ever Work, Not On-time	Never Work, Not On-time
	First Stage				
IV, Minority	0.4469***	0.0053	-0.0266**	0.0345***	-0.0130*
	(0.0063)	(0.0151)	(0.0119)	(0.0109)	(0.0075)
IV, White	0.3643***	-0.021	0.0422	-0.0388	0.0259
	(0.0201)	(0.0619)	(0.0563)	(0.0335)	(0.0251)
P-value, test of diff.	0.00	0.681	0.233	0.038	0.139
CCM, Minority		0.721	0.122	0.115	0.041
CCM, White		0.791	0.097	0.115	0
		Panel B: Any Graduation			
		Ever Work, Graduated	Never Work, Graduated	Ever Work, Not Graduated	Never Work, Not Graduated
IV, Minority		0.0276*	-0.0322***	0.0109	-0.0063
		(0.0150)	(0.0122)	(0.0097)	(0.0068)
IV, White		-0.0258	0.0492	-0.0477*	0.0175
		(0.0615)	(0.0570)	(0.0290)	(0.0234)
P-value, test of diff.		0.399	0.163	0.056	0.331
CCM, Minority		0.753	0.131	0.084	0.032
CCM, White		0.811	0.09	0.11	0
		Panel C: Any Graduation or Continued Attendance			
		Ever Work, Persisted	Never Work, Persisted	Ever Work, Not Persisted	Never Work, Not Persisted
IV, Minority		0.019	-0.0251**	0.0199**	-0.0131**
		(0.0150)	(0.0126)	(0.0090)	(0.0062)
IV, White		-0.0307	0.0436	-0.0417	0.0221
		(0.0629)	(0.0583)	(0.0286)	(0.0213)
P-value, test of diff.		0.443	0.25	0.040	0.113
CCM, Minority		0.788	0.132	0.049	0.030
CCM, White		0.849	0.1	0.07	0

Notes: N = 12,589 Minority youth and N = 1,085 White youth. 58 observations in graduation data are dropped due to missing race/ethnicity. Sample and outcomes defined in Table 6. CCM shows control complier means, rounded to 0 if estimate is negative. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Figure 1: Experimental Flow Chart

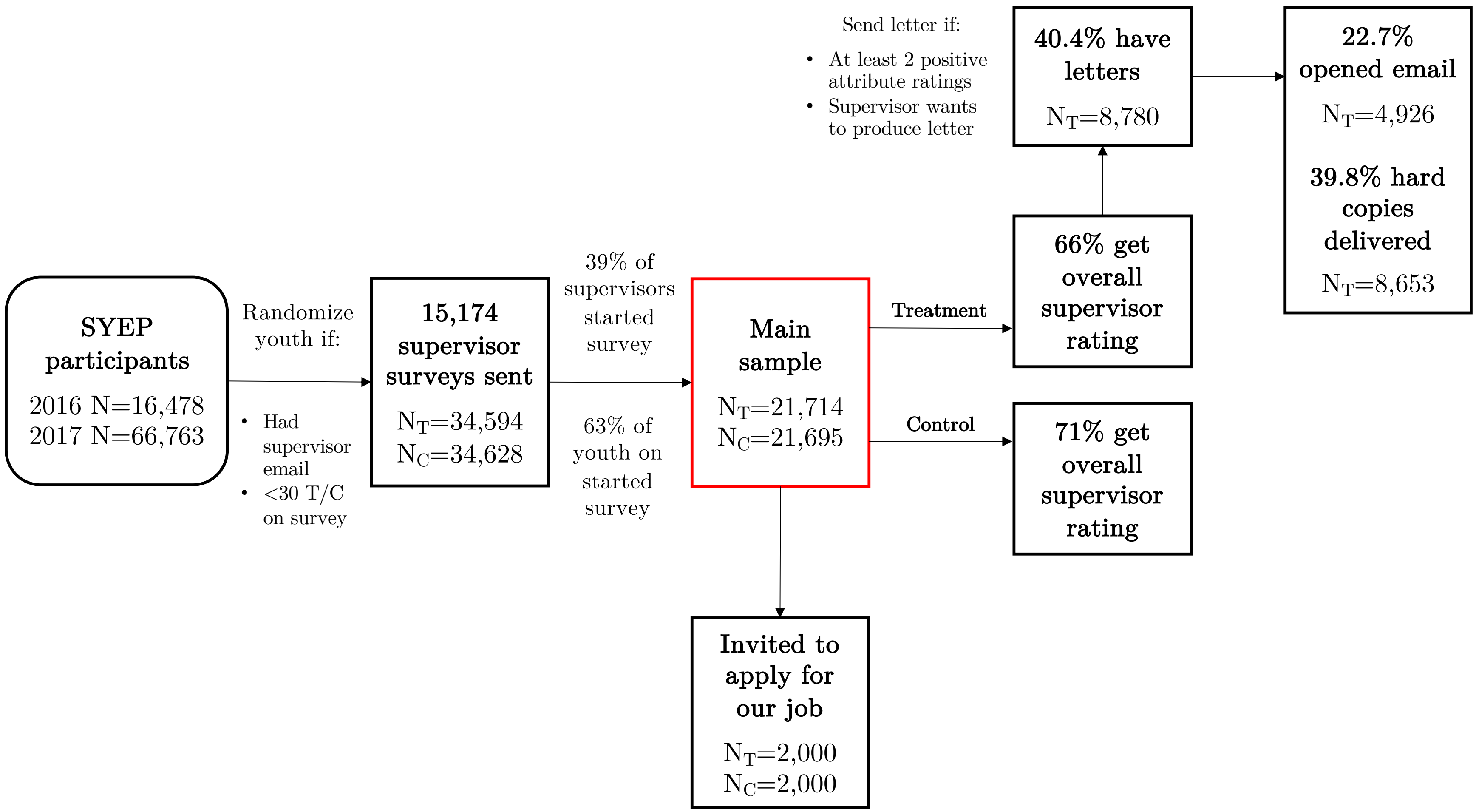
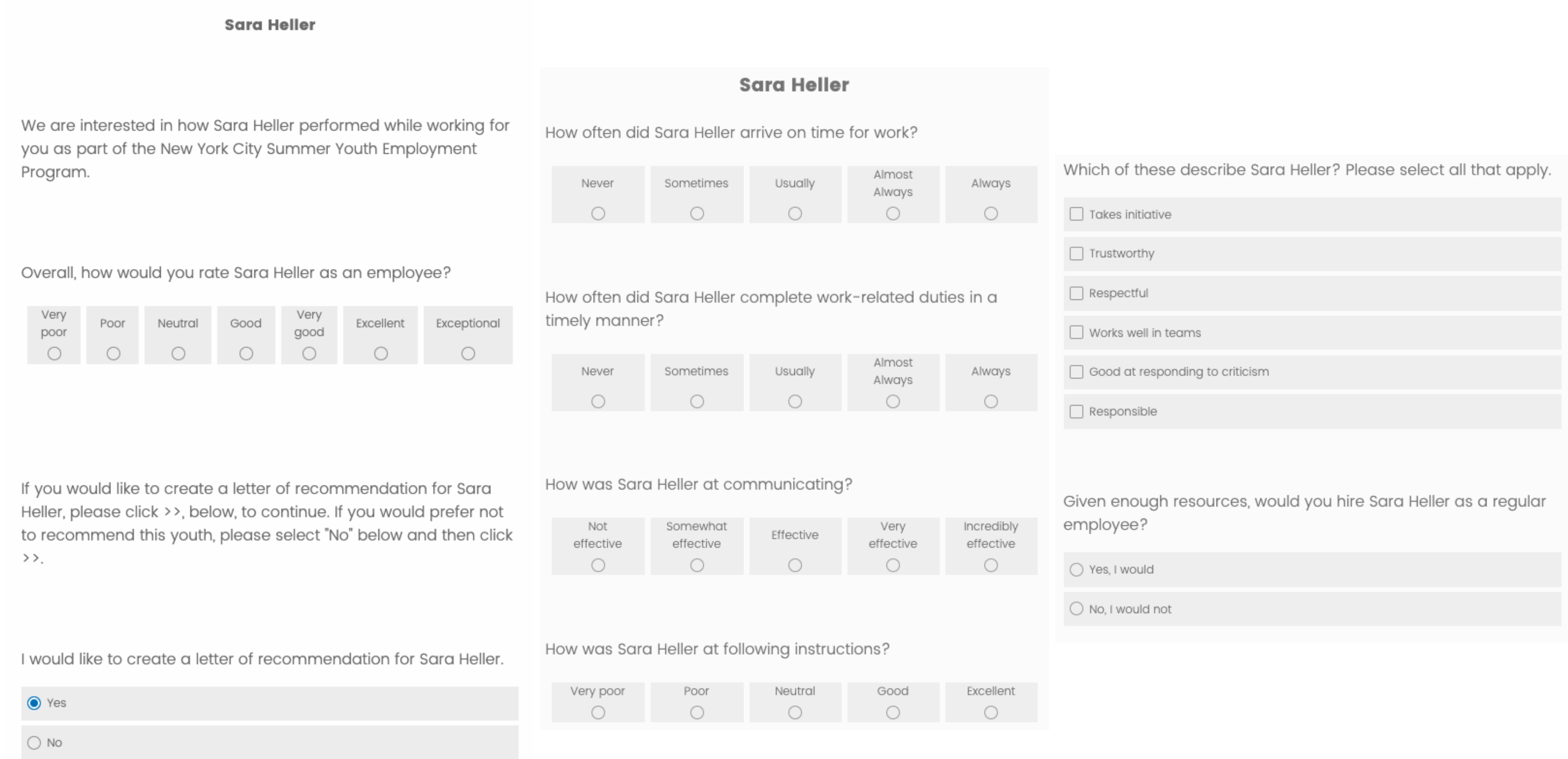


Figure 2: Screenshots about Treatment Youth on Supervisor Survey



Notes: The image on the left shows the first screen supervisors saw asking about each youth with the overall rating question and the invitation to write a letter. As indicated in the image, the option to create a recommendation was pre-selected. The images in the middle and on the right show the questions asked about each treatment youth when the supervisor agreed to create a letter of recommendation.

Figure 3: Example Letter of Recommendation

Department of
Youth & Community
Development

November 1, 2017

To Whom It May Concern:

Sara Heller worked for me at the Wharton School during the summer of 2017. Overall, Sara was an exceptional employee.

With regard to reliability, Sara was always on time to work. Sara always completed work related tasks in a timely manner.

When it came to interpersonal interaction, Sara was an incredibly effective communicator. Sara was excellent at following instructions.

In addition to Sara's other strengths, Sara takes initiative, is trustworthy, is respectful, works well in teams, is good at responding to constructive criticism, and is responsible.

Given the resources, I would hire Sara as a full-time employee. I invite you to contact me if you would like more information. I can be reached at 215-898-7696 or judd.kessler@wharton.upenn.edu.

Sincerely,

Judd Kessler
The Wharton School

The New York City Department of Youth and Community Development (DYCD) invests in a network of community-based organizations and programs to alleviate the effects of poverty and to provide opportunities for New Yorkers and communities to flourish.

Empowering Individuals • Strengthening Families • Investing in Communities

Note: This recommendation letter is part of a pilot program being run by the New York City Department of Youth and Community Development. Some youth were randomly selected to be part of the pilot. These youth were eligible to receive a letter of recommendation, which reflects supervisor feedback about each individual's job performance.

Notes: See Figure 2 for the source of inputs into each sentence for this example letter.

ONLINE APPENDIX

The Effects of Letters of Recommendation in the Youth Labor Market

Sara B. Heller
University of Michigan & NBER

Judd B. Kessler
The Wharton School & NBER

A.1 Additional Labor Market Results, Main Sample

A.1.1 Employment and Earnings

The main text reports annual and cumulative earnings results for two functional forms of the earnings variable: winsorized at the 99.5th percentile and log(earnings+1). Table A.1 shows other transformations of the raw dollar amounts, including an alternative winsorization (at the 99th percentile), alternative intercepts added to earnings prior to logging (0.1, 10, and 100), and the inverse hyperbolic sine transformation. The alternative winsorization in Panel A makes very little difference relative to the results in the main text. The other panels show that, as expected given that there are treatment effects on the extensive margin, the decision about what to add to the 0s does change the point estimate somewhat. Cumulative earnings increases for compliers range from 9.5 to 23 percent, all significant at the 5 percent level. The results suggest that although there is some uncertainty about how big the proportional change in earnings is, the basic conclusion of a significant increase in earnings is robust to different functional form choices. We emphasize the 4 percent earnings increase from the winsorized version in the main text both because we pre-specified winzorized earnings as our primary outcome, and because it is the most conservative result. However, we also pre-specified that we would check alternative functional forms given the likely importance of skewness in the data.

Some additional nuance in the pattern of labor market results is shown in Figure A.7, which breaks down both the employment and earnings results by quarter rather than by year. Although the quarterly results are fairly imprecise, two general patterns are of interest. First, we would expect the cumulative employment effect to go to zero eventually: almost everyone in the sample is likely to work at some point, leaving no more room for letters to change whether someone ever worked. But as Panel A shows, it still appears that treatment youth work more in each quarter when measured non-cumulatively; in every quarter the point estimate is positive. Second, the effects are not concentrated during the summer quarters (quarters 3 and 7, as indicated by the much higher control means listed at the bottom of the graph). Rather, it appears that the letters increase employment throughout the year. Similarly, Panel B shows that the earnings effects also do not fade out over time. Point estimates are increasing in dollar amounts, while remaining proportionally similar relative to the growing control means.

A.1.2 Spell Length

The main text argues that the mechanism underlying the employment improvements involves employers updating their beliefs. If employers' previous experience is that only extraordinary employees include letters of recommendation in their applications, there could be a risk that employers inefficiently update (e.g., believing that treated applicants will be more productive workers than they actually being). Table 3 in the main text shows that the recommendation letters increased the number of quarters worked and decreased the time until the first quarter worked without increasing the number of job spells, which implies that youth must be working longer in their jobs.

Table A.2 directly confirms that youth are working longer without switching jobs, providing further evidence that the letters are not generating worse matches. Each panel shows results for a different job spell, with spell 1 being the spell started the earliest, spell 2 being the spell started next, and so forth. If spells are started in the same quarter, we assign the longer spell the lower spell number. We count any spell with at least one quarter occurring in the post-letter period. Youth must have a given spell number to appear in each panel, so the sample becomes more selected as the spell number rises. The first column reports treatment effects on the length of each spell, defined as the number of consecutive quarters worked at the same employer. The treatment effect on the length of the first and second spells is positive, but only statistically significant for the second spell, which increases by 0.11 quarters (about 6 percent) for letter recipients.

Because part of the treatment effect is to help youth find jobs more quickly, it is possible that total spell length is biased from differential censoring at the time our data end; that is, we may observe treatment spells for more quarters after their start than we do for control spells. The second column tests this possibility by using an indicator for whether a spell is censored by the end of our data (i.e., whether a youth was still working at an employer in the last quarter we have in the data). Point estimates are all substantively quite small and not statistically significant, suggesting that differential censoring is not biasing our spell length results; spells are frequently short enough that we observe the entire spell, despite the earlier start for treatment youth. The last 3 columns of the table confirm that the results are robust to looking only at spells that are not censored. We report treatment effects on whether a spell lasts at least 2, 3, or 4 quarters, conditional on observing all the quarters. Overall, there is no evidence that letters are creating bad matches. Rather, they appear to be generating longer job durations.

A.1.3 Employer Type

Tables A.3 and A.4 separate employment and earnings effects by the type of employer. Because the letter came on DYCD letterhead (the agency that runs the SYEP), it is possible that the letter increased the rate at which youth reapplied to or engaged with future SYEP activities or other term-time work where DYCD was the employer of record.

Table A.3 shows that this is not a main driver of our results by reporting labor market results separately for DYCD and for all other employers. The only significant increase in employment is at non-DYCD employers, meaning that the letters increased employment outside the SYEP agency. Earnings impacts are directionally much larger at non-DYCD employers, on the order of 5 rather than 0.5 percent, although estimates are too imprecise to draw strong conclusions.

Table A.4 explores what types of industries letter recipients work more in. The classification across industry clusters is based on Gelber, Isen, and Kessler (2016), which groups industries that are over-represented in SYEP, like childcare and landscaping (cluster 1) separately from industries that are under-represented in SYEP, such as retail and food service (cluster 2). Directionally, letters seem to increase employment in both types of industries, with earnings increases concentrated in cluster 2 jobs. This pattern suggests that the letters are helping young people find jobs even outside of the industries that they were most likely to be exposed to through SYEP.

A.2 Additional Job Application Results

Table A.5 shows that observable baseline characteristics are balanced for the randomly selected subsample to whom we sent our job advertisement. No more differences are significant than we would expect by chance, and the joint F-test fails to reject the null hypothesis that the groups are the same ($p = 0.215$). Notice, however, that there is some chance imbalance on the proportion of the subsample that is Hispanic, with 32.8 percent of the treatment group and 29 percent of the control youth in this category ($p = 0.01$).

As we show below, labor market impacts are biggest for Hispanic youth. As a result, the chance imbalance on ethnicity means that the labor market effects are somewhat larger in our job application subsample than in our main sample, despite being a randomly selected subset. Employment effects are similar to the main sample, but the cumulative earnings IV estimate is \$1870 ($se = 956$, $p < 0.1$). Consistent with the argument in the main text that the increase in employment is what slows down progress in school, point estimates on education outcomes are also slightly more negative in the job application subsample, though still not distinguishable from zero.

A.3 Differences by Supervisor Ratings

As discussed in the main text, we designed the survey to maximize the information we would have available to produce recommendation letters, not to ensure that treatment and control youth would be treated equally on the survey. As such, we asked about each treatment youth first, on the same page as we asked supervisors to decide whether to produce a letter. After the supervisor had seen all treated youth, we then asked them a single question about the overall performance of each control youth on the same page. This aspect of our design makes it likely that supervisors would use different decisions rules when assessing whether to give a particular youth a rating.

Indeed, treatment youth are significantly less likely to have been rated by a supervisor (66 versus 71 percent had a rating, $p<0.01$). Despite the potential for selection into having a rating, the other observable characteristics are generally still balanced, with a joint F-test failing to reject equality across all observables ($p = 0.605$). Table A.6, however, which breaks out the balance tests for youth receiving low versus high ratings, shows that there is some imbalance within the group that receives low ratings ($p = 0.094$).

Because of the dramatic difference in having a rating and the small imbalance on observables for those with low ratings, we focus on the subsample of rated youth on complete surveys in the main text. But Table A.7 shows that the pattern of heterogeneity in labor market impacts across rating categories is quite similar to those shown in the main text, even when all youth with an employer rating are included.

Table A.8 shows the equivalent balance tests for the subsample of youth who appeared on a fully completed survey (i.e., where the employer rated every youth on the survey). Although this is a selected group, full survey completion limits the scope for treatment and control youth to be differentially selected into getting a rating. Indeed, the difference in receiving a rating is much smaller in this sample: 31.6 percent for treatment youth and 32.5 percent for controls ($p = 0.066$). But as the table shows, observables are entirely balanced within each rating group. As a result, this is the subsample we use to assess how treatment effects vary by rating in the main text.

We have additionally tested whether treatment effects on applying to our job posting are different for those with a high versus low rating. Given that this limits an already reduced sample (N = 4,000) to those with ratings (N = 2,783), and then splits the sample into groups, this is not a highly powered test. The difference in the intent-to-treat effects for the high-rated group relative to the low-rated group (i.e., the interaction effect between treatment and being highly rated) is $\beta = 0.008$, $p = 0.721$, with a control rate of application for the low group of 0.078. The difference for the IV is $\beta = 0.018$, $p = 0.748$. So while it is possible that receiving a letter had a more positive effect on job search behavior for highly-rated youth, we cannot reject the null that both effects were zero.

A.4 Education Analysis

A.4.1 Sample Definitions

Because we wrote our pre-analysis plan prior to our conversations with DOE about what data would be available, and prior to matching to our study sample to assess data coverage, our education results are where we deviate most from our pre-analysis plan. We initially expected to test an index that included days present, an indicator for graduating or still being in school, GPA, and standardized test scores when available, plus a separate outcome measuring post-secondary enrollment. In practice, many elements of this index are missing for multiple reasons. Many students are not in school to have attendance, or they attend a school (including charters) where DOE does not share records; we do not have standardized test scores in the data (other than the selected group that takes Regents exams); and DOE measures graduation and college enrollment only for particular cohorts at particular times.

So instead of forcing different patterns of missing outcomes into a single index, we instead present results for a similar set of outcomes, but separately by outcome for groups defined by data availability. In particular, we define three samples for the education analysis: those expected in high school records, in graduation records, and in college records.

We wish to avoid missing data from students who had already left school, transferred, or attended charter schools. But we cannot define our sample based on whether they have schooling records during outcome years, since treatment could affect enrollment. Instead, we define our high school sample using only baseline characteristics. We identify students who were in public, non-charter schools, attending grades 8–12 in the pre-randomization year, but who had not graduated by the August prior to the academic year the study took place. This is the group we would expect to see in high school records if they progressed through high school without transferring or dropping out (we explain how we handle missing data for each outcome in the next section). This "expected in high school" sample leaves us with 19,714 students in our main education analysis, with treatment and control youth equally likely to appear ($\beta = -0.002, p = 0.676$).

To define our main graduation sample, we limit the "expected in high school" sample to those who, prior to randomization, were enrolled in a grade where we could observe most or all of their graduation outcomes prior to the end of our graduation records (i.e., the 2019–20 academic year). This leads us to only include youth who were in 10th–12th grade in the year prior to randomization. Figure A.8 shows the group that comprises our main 10th–12th grade graduation sample in gray. It excludes those in grades that were too young to have any 5th- and 6th-year graduation information (as well as 12th graders who already graduated by the time letters were distributed). Since we can observe some, but less complete, graduation information for the younger cohorts as well, we show graduation results including those cohorts as a robustness check. There is no treatment-control difference in the probability of appearing in either our main graduation sample or the sample including younger cohorts.[1a]

DOE captures post-secondary enrollment data at a single point in time, 6 months after a student reaches their on-time graduation date (i.e., only on-time graduates have college enrollment recorded in the data). This information is based on data from the National Student Clearinghouse and from the City University of New York. Because of the timing of this measure, our post-secondary enrollment analysis makes one additional limitation relative to the graduation sample: it also excludes all pre-randomization 12th-graders from the "college analysis" sample, since their on-time graduation date makes their college outcome a baseline characteristic (measured just before our letters were distributed). We also have treatment-control balance on the probability of being in this sample ($\beta = -0.002, p = 0.756$).

1a. Note that the graduating cohort is defined by the official 9th grade cohort to which a student belongs per state standards. We do not directly observe which graduation cohort a student is in if they are not in our graduation records, so we limit the sample sample based on whose pre-randomization grade puts them in a graduating cohort for which we would have an observed outcome, if the student were to graduate on time relative to their pre-randomization grade. That means that even students who transferred to other districts during the outcome period will remain in our data; we discuss their outcome definition below.

A.4.2 Outcomes

A.4.2.1 School Performance

Because missing data grows over time as students graduate or drop out, we focus our high school outcome analysis on the academic year of random assignment (letters were distributed in November or December). To measure high school engagement, we use an indicator for any enrollment, the number of days enrolled, the share of enrolled days actually attended, and GPA on a 100-point scale (although some students have over 100 due to the up-weighting of advanced classes). We assign 0 days present and 0 days enrolled to students in the expected in high school sample who are missing after treatment, though they may have attended school outside our data coverage. We analyze non-missing GPAs only. Since there is treatment-control balance on whether someone is in the enrollment, attendance, and GPA data, alternative imputations of missing data would not change results.

A.4.2.2 Graduation

To measure on-time graduation, we create an indicator for whether a student in the graduation sample earned a diploma (local, Regents, or Advanced Regents) by her 4th year after initially entering 9th grade.[2a] To measure any graduation, we create an indicator for whether a student earned a diploma at any point during our follow-up period. Note that the data only include graduation information if a student's 4th, 5th, or 6th year for graduation falls between Fall 2015 and Summer 2020 (see Figure A.8). We assign anyone who is part of the graduation sample but missing graduation information a 0 for not receiving a diploma from the NYC DOE. As we report in the main text, there is no treatment effect on transferring, so even if we assumed that all transfers graduated elsewhere, it would not change the results. We also define a final indicator to measure school persistence, defined as 1 if someone has either graduated or has non-zero days present in the final year of our data (2019–20). This will help to capture those who are still working toward a degree but run into the end of the data before their 6th year for graduation, as well as those who keep attending school after their 6th-year graduation date.

A.4.2.3 Post-Secondary Enrollment

We measure any post-secondary enrollment as whether someone is enrolled in a 2- or 4-year institution 6 months after what would have been their on-time graduation date (which is the only timing available in the data). We do not count participation in vocational or public service post-secondary activities as college enrollment. As with graduation, we assign a 0

2a. As described above, the 9th grade cohort is determined by the state of New York, so cohorts are defined in the graduation records even when we do not observe the initial 9th grade year in our data.

from anyone who is part of the college sample but missing from the post-secondary data.

A.4.3 Descriptive Statistics

Table A.9 shows descriptive statistics for two groups: our "expected in high school" sample in the left panel, contributing to the broader education analysis, and our "graduation sample" in the right panel, the focus of our joint outcome analysis in the main text. On average, students in our education sample are about 16 years old, 45 percent male, 42 percent Black, 31 percent Hispanic, 14 percent Asian, and 8 percent White. They are in 10th grade on average, attending about 90 percent of the days they are enrolled, and earning a C-plus average. Over 60 percent of them had not worked in UI-covered jobs prior to the SYEP. The table also shows that across all baseline characteristics, treatment and control groups are jointly balanced ($p = 0.151$). It is worth noting that there is some chance imbalance on GPA and on the proportion of the sample that is White; although the differences are substantively small (-0.39 on a 100-point GPA scale and 1 percentage point more likely to be White), they are statistically significant. As a result, the exact magnitude of the education results are slightly more sensitive to how covariates are included in the regressions (see Appendix Section A.9). However, none of the substantive conclusions change regardless of covariate choice.

Similarly, there appears to be some chance imbalance among the 10th–12th graders who make up our main graduation sample, both on several individual characteristics and jointly (p=0.073). The treatment and control means show that the differences tend to be quite substantively small (e.g., the treatment group averages 0.5 fewer GPA points on a 100-point scale and is 1 percentage point more White). As noted in the main text, one benefit of the post-double-selection method of covariate selection is that we can control for these chance differences without specification mining.

A.4.4 Employment Results by Inclusion in the Education Sample

Table A.10 compares labor market impacts for those who are and are not in the expected in high school sample. Both groups respond positively to the letters. The employment effects are slightly larger for those in the education sample, though earnings impacts are slightly smaller. None of the differences between these groups is statistically significant.

A.4.5 Treatment Effects on Education Outcomes

Table A.11 shows the ITT and IV impacts of the letters of recommendation on education outcomes, first for our expected in high school sample in Panel A, then, for completeness, on the subset of youth who are in our graduation sample in Panel B. For the whole expected in high school sample, the first four columns show that, in the academic year we send letters, there are no significant changes in enrollment, days enrolled, the share of days enrolled actually attended, or GPA. Point estimates are relatively small, with confidence intervals ruling out treatment effects more than 1 percent in either direction.

We explore longer-term educational outcomes in the other columns. While none of the treatment effects is statistically significant, we highlight one pattern that becomes significant for the non-White subgroup and in alternative covariate specifications: on-time (4-year) graduation shows a substantively important decline, while the point estimates on ever graduating (including delayed graduation), graduating or still attending, and enrolling in college on time, are much closer to zero. The fact that on-time college attendance is not declining in the same way that on-time graduation is suggests that the slow-down in high school completion is among the students who are not going directly to college.

We conclude that there is little evidence that letters improve student performance in school (e.g., by changing teacher or guidance counselor beliefs or encouraging college application). They may, however, slow down progress towards a diploma, likely by pulling marginal students out of school and into the labor force, as discussed in the main text.

Panel B limits shows these same effects for the subset of youth in our graduation analysis. Point estimates are somewhat more negative, but still generally insignificant, with the exception of a marginally significant decline in on-time graduation. This is consistent with the main results on the joint graduation-employment outcomes in Table 6.

Table A.12 shows the same minority-White breakdown for education outcomes as the main text shows for labor market outcomes, focusing on our expected in high school sample. The decline in on-time graduation is entirely concentrated among non-White high school students, for whom there is a marginally significant 1.7 percentage point (2 percent) decline.[3a] The fact that the negative effect of the letters on education outcomes are concentrated in the same group that sees a positive effect of the letters on labor market outcomes further corroborates the hypothesis that the increase in labor market engagement is generating some decrease in school engagement.

A.5 Robustness Checks for Joint Outcomes

To streamline the presentation, the main text presents only the IV results for the main outcomes combining graduation and employment information. Tables A.13 and A.14 show the ITT versions overall and by race/ethnicity.

The main text mentions that the shift from graduating on time without working to working without graduating on-time is concentrated among those with below-median GPAs.

3a. The IV estimate just barely crosses the $p = 0.1$ threshold, differing from the marginally significant ITT due to small differences from covariate adjustment.

These results are shown in Tables A.15 (ITT) and A.16 (IV). We define the median based on non-missing, one-year (i.e., not cumulative) GPAs from the pre-randomization academic year, which is a B-minus (81.64). Both the shift towards work and the shift away from schooling is concentrated among below-median students, and the groups are sometimes statistically distinguishable, depending on the outcome. If we instead break these groups down by quartile of GPA, the results are only significant in the lowest quartile, although results suggest that the letters may affect the second quartile as well.

Finally, we note that we constructed the graduation sample to include mostly students who have had time for 4th-, 5th-, and 6th-year graduation to be fully observed (with the exception of one cohort of 10th graders, for whom the data ends after their 5th year). This choice helps ease interpretation, since the majority of that sample have had time for all graduation outcomes appear in the data. However, we can also observe a subset of graduation information, (e.g., on-time but not later graduation), for youth as young as 8th grade in the pre-randomization year (see Figure A.8). For completeness, we show the joint outcome results for the 8th–12th graders in our expected high school sample as well (ITT in Table A.17 and IV in Table A.18). As would be expected given that not everyone has had time to realize their full graduation outcomes, the only significant changes are for the on-time and school persistence measures. The rest of the point estimates largely show the same pattern as the main graduation sample, but with smaller and noisier point estimates. Once a couple additional years pass, we should be better able to measure graduation outcomes for the whole education sample.

A.6 Heterogeneity

A.6.1 Pre-Specified Categories: Race/Ethnicity, Gender, High School, Age, and Neighborhood

Tables A.19 through A.27 show treatment effects for different subgroups of youth. Because of the number of hypothesis tests across these tables, and the limited statistical power, we do not emphasize the statistical significance of any particular result. However, we pre-specified an interest in these divisions as exploratory, so we attend to the basic patterns here.

Tables A.19 and A.20 show additional labor market results separately for the same division between White and non-White youth as in the main text. Table A.19 shows the different earnings skewness transformations, and Table A.20 shows impacts on the amount of time and number of spells worked. Consistent with the main results on race/ethnicity, all of these results show that the main labor market effects are concentrated among minority youth.

Tables A.21 and A.22 further break down the main labor market results separately by race and ethnicity subcategories (ITT and IV respectively). They show that employment impact is driven by somewhat larger effects for Asian and Hispanic youth, with earnings effects suggestively larger for Hispanic youth. Both groups are more likely to get a letter (the first stage in the second panel of each table is larger), but even among compliers the effects are larger for these minority youth. However, as in the main results, we are underpowered to detect group differences; we cannot reject the null that effects are the same across all groups.

A similar pattern holds for women (Table A.23), with female SYEP participants significantly more likely to receive a letter, and with compliers having suggestively larger employment effects despite similar or smaller cumulative earnings effects. This finding is consistent with the Abel, Burger, and Piraino (2020) result that the employment benefits of recommendation letters in South Africa were concentrated among women. But unlike in that setting, young women in NYC do not face the same difficulty finding work relative to young men; indeed, consistent with broader U.S. patterns of young women outperforming their male counterparts, employment rates for women are considerably higher than for men in our sample. The fact that there are larger effects for women both in settings where priors are likely to favor and to disfavor women suggests that the effect is not simply about statistical discrimination, since priors should go in the opposite direction across settings.

Table A.24 shows effects separately for youth who report still being in high school or not being in high school at the time of application to SYEP. Note that this is different than the sample that is expected to be in high school within the education data, since here we use applicants' self-reports, so that we have an education status for everyone, regardless of enrollment in NYC's public school system. Employment effects are suggestively bigger for those still in high school at the time of SYEP application, although earnings effects are more similar across the groups.

Table A.25 shows effects for those under 18 and those 18 and over at the time of application. Employment point estimates are slightly larger and earnings estimates slightly smaller for those under 18. But both sets of effects are statistically indistinguishable from the effects for older youth.

Table A.26 shows effects by neighborhood economic mobility. Using the Opportunity Insights "upward mobility" data, we use each individual's zip code to assign their neighborhood an average income rank for children whose parents were in the 25th percentile of the national household income distribution. Opportunity Insights provides these data at the Census Tract level. We use the Zip Code Tabulation Area (ZCTA) crosswalk to map Census Tracts onto zip codes, which is the geographic information we have on our sample. In cases of multiple Census Tracts falling within a given ZCTA, we use the average upward mobility value (i.e., the unweighted mean across all upward mobility values that fall within the ZCTA). We divide the youth into those who live in areas with above and below median mobility, with median defined in-sample. Table A.26 shows labor market impacts for these two groups. There are positive effects for both both those living in above- and below-median neighborhoods, with employment and earnings impacts suggestively larger in places with below-median mobility.

A.6.2 Previous Work Experience

Table A.27 shows labor market impacts separately for young people who did or did not have any prior work experience (measured as appearing in the UI data) before the SYEP summer. Although this was not a pre-specified subgroup of interest, we show these results as an exploration into potential mechanisms. In theory, if statistical discrimination is the sole driver of program impacts, we might expect to see bigger effects for the group with more uncertainty about their productivity (i.e., those without other work histories).

In fact, we see the opposite: point estimates are larger and only statistically significantly different from zero for the group that had previous work experience. As with the gender heterogeneity in the previous section and with the rating heterogeneity in the main text, this result seems to be more consistent with the possibility that employers are using the letters to help identify those likely to be higher performers, rather than to just improve their priors about those with the least available information.

A.7 Information on Letters by Subgroup

To help interpret the patterns of results by subgroup, Table A.28 shows some additional information about the letters for the different subgroups discussed in the previous sections. The entire table shows the treatment group only, since they were the only ones eligible for a letter. The first column shows the proportion of each group that was sent a letter (i.e., having a supervisor agree to produce one and receiving ratings high enough to generate a latter); this summarizes the information shown in the "first stage" column of the separate heterogeneity results. The second column is conditional on the first, showing average overall employee rating on a scale from 1–7 for those who were sent a letter. The third column shows the proportion of each group that submitted an application in response to our job application, conditional on being one of the 2,000 treatment youth randomly selected to receive the job advertisement. The fourth column, conditional on the third, shows the proportion of the applicants that uploaded a letter of recommendation (ours or any other) as part of their application.

There is significant variation both in letter receipt and in average ratings. Non-white, female, non-high school, previously-employed, and below-median neighborhood mobility youth are all more likely to receive a letter. But the higher rate of letter receipt does not always correspond with stronger letters, on average. To focus on the minority-White difference in the main text, minorities actually have significantly lower average ratings conditional on receiving a letter than their White counterparts. And they do not use the letter more frequently; their rate of letter usage is actually about 6 percentage points lower than the White youth who applied to our job posting, although the small sample size limits how well we can differentiate the groups. The basic pattern of results suggests that the larger effects for minorities are likely to be driven by how employers respond, even to slightly weaker letters, rather than big differences in how the groups use the letters.

The only significant differences in letter usage are between those who were or were not in high school at the time of SYEP application, and relatedly, those who were under 18 versus 18 and older. This likely helps to explain the bigger labor market point estimates for high school youth, who were much more likely to use the letter on our job application than those who were not in school.

A.8 Comparing Our Main Sample and Everyone on a Survey

The main text focuses on the sample of youth who were on a survey that a supervisor started, a group that we pre-specified as of special interest in our pre-analysis plan. This excludes 25,813 young people who were only on surveys that no one started. Since none of these individuals could possibly have been treated if assigned to treatment, everyone in this group is effectively a never-taker. Since we are able to observe this fact for both treatment and control youth on these surveys, we exclude them from our main analysis.

This section provides some additional information on who is excluded from the sample and the implications for our analysis. Table A.29 compares our main control group to everyone who was on an unopened survey (treatment and control) on baseline characteristics and main outcome measures.

Given that assignment to supervisors was not random, it is not surprising that young people whose supervisors did not start the survey are observably different than those in our main sample. Table A.29 shows that our main sample is younger, less Black and less White (more Hispanic and Asian), more likely to still be in high school, and generally less engaged in the labor force pre-randomization than those on unopened surveys. Table A.30 shows that our control group continues to be less involved in the labor market than those on unopened surveys during the outcome period, but more engaged and successful in school. While not shown in the table, we find no significant difference in job application behavior, consistent with the argument in the main text that employment status does not affect the decision of whether to apply to our job.

Given the observable differences between our main sample and those on unopened surveys, our estimates are most externally valid for the group that would look most like those in our main sample: young people whose supervisors fill out the surveys when asked, without any requirement to do so. It is possible that forcing supervisors to fill out surveys for their employees could generate somewhat different effects, given that the population of youth affected would be observably different.

Tables A.31 and A.32 show our main results without excluding those on unopened surveys. As expected given that this adds solely untreated individuals regardless of treatment status (i.e., massively increases the non-compliance in the sample), effects are uniformly smaller and less significant. The main sample's point estimates are within the confidence intervals of these point estimates, also consistent with the fact that the inclusion of unopened surveys just adds noise.

A.9 Robustness to Different Covariate Choices

The main text uses the post-double-selection LASSO (Belloni, Chernozhukov, and Hansen 2014a, 2014b; Belloni et al. 2012) to choose which covariates are included in each regression. For robustness, this section shows two different alternatives: including no covariates other than the cohort indicator needed for treatment to be conditionally random (i.e., controlling for randomization strata), and including all covariates that we feed into the post-double-selection process.

For employment outcomes, the covariates we feed into the lasso include indicators for: being male; being employed in each of the 2nd through 6th years prior to randomization; the earnings quartile of the pre-randomization year earnings; never being employed pre-SYEP; self-reporting being in high school, college, or being a high school graduate; being 15–16, 17–18, 19–20, or 21 and older; being part of the Ladders for Leaders program; being Hispanic, Asian, White, Other, or having missing race/ethnicity; not being matched to the education data; and being in the expected in high school sample. For the education outcomes, covariates we feed into the lasso include indicators for: being in grade 8 or under, grade 10, grade 11, or grade 12; being in deciles 1 through 9 of prior year GPA or missing GPA; being in quartiles 2 through 4 of the share of enrolled days attended; being male; being employed in each of the 2nd through 6th years prior to randomization; the earnings quartile of the pre-randomization year earnings; never being employed pre-SYEP; self-reporting being in high school, college, or being a high school graduate; being 15–16, 17–18, 19–20, or 21 and older; being part of the Ladders for Leaders program; and being Hispanic, Asian, White, Other, or having missing race/ethnicity. In regressions using the graduation sample, the quartile indicators are calculated separately for the graduation sample, which just covers 10th–12th graders.

Tables A.33 and A.34 show alternative results for labor market and education effects respectively, using either no covariates other than the randomization stratum indicator needed for conditional independence, or using all covariates. These tables lead to the same conclusions as the main tables. Because of the imbalance in several education baseline covariates discussed in section A.4, the point estimates on GPA and on-time graduation become somewhat larger and more significant with these different covariates specifications are used instead of the lasso. It is for partly this reason that we take the decline in on-time graduation seriously as a main result, even though it is not statistically significant on average in our main results.

A.10 Additional Figures and Tables

Figure A.1: Example Supervisor Survey Invitation Email

Dear Judd Kessler,

Thank you for your participation in the 2017 Summer Youth Employment Program (SYEP), run by the New York City Department of Youth and Community Development.

For the second year, we are running a "letter of recommendation" program. As part of this program, **we are asking you to complete a very short survey** about some of the youth who worked for you this summer (the survey should take about 1 minute per selected youth).

Positive responses will be turned into letters of recommendation for the youth. We expect these letters to help youth capitalize on their experience working for you this summer.

To join employers like you in participating, please click on this personalized link by **a week from tomorrow, Friday, October 20th**: Take the survey.

If you have any questions about the program, please see a further description on our website here.

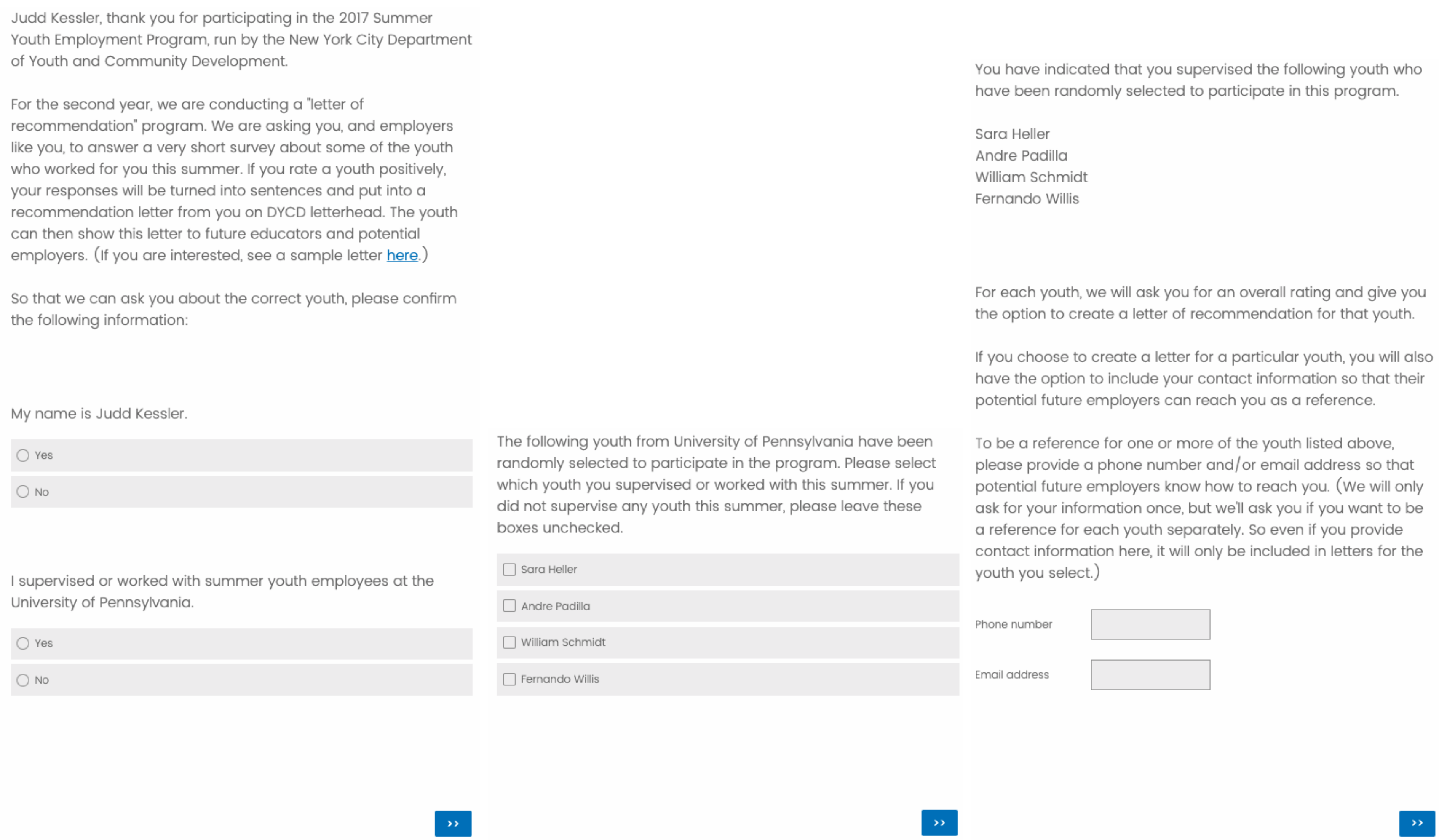
If you have additional questions, you can contact our academic partners: Judd B. Kessler (judd.kessler@wharton.upenn.edu) at the University of Pennsylvania and Sara Heller (hellersa@sas.upenn.edu).

Sincerely,

SYEP Team

Follow the link to opt out of future emails:
Click here to unsubscribe

Figure A.2: Screen Shots from Beginning of Supervisor Survey



Judd Kessler, thank you for participating in the 2017 Summer Youth Employment Program, run by the New York City Department of Youth and Community Development.

For the second year, we are conducting a "letter of recommendation" program. We are asking you, and employers like you, to answer a very short survey about some of the youth who worked for you this summer. If you rate a youth positively, your responses will be turned into sentences and put into a recommendation letter from you on DYCD letterhead. The youth can then show this letter to future educators and potential employers. (If you are interested, see a sample letter here.)

So that we can ask you about the correct youth, please confirm the following information:

My name is Judd Kessler.

- Yes
- No

I supervised or worked with summer youth employees at the University of Pennsylvania.

- Yes
- No

››

The following youth from University of Pennsylvania have been randomly selected to participate in the program. Please select which youth you supervised or worked with this summer. If you did not supervise any youth this summer, please leave these boxes unchecked.

- Sara Heller
- Andre Padilla
- William Schmidt
- Fernando Willis

››

You have indicated that you supervised the following youth who have been randomly selected to participate in this program.

Sara Heller
Andre Padilla
William Schmidt
Fernando Willis

For each youth, we will ask you for an overall rating and give you the option to create a letter of recommendation for that youth.

If you choose to create a letter for a particular youth, you will also have the option to include your contact information so that their potential future employers can reach you as a reference.

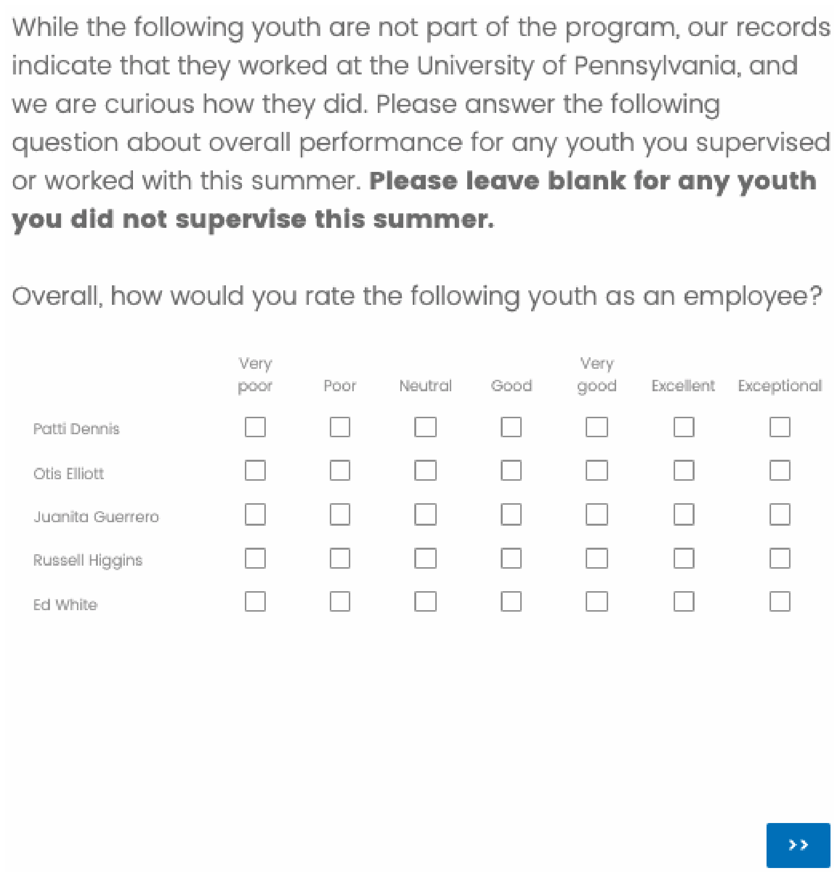
To be a reference for one or more of the youth listed above, please provide a phone number and/or email address so that potential future employers know how to reach you. (We will only ask for your information once, but we'll ask you if you want to be a reference for each youth separately. So even if you provide contact information here, it will only be included in letters for the youth you select.)

Phone number

Email address

››

Figure A.3: Screenshot of Control Youth Rating on Supervisor Survey



While the following youth are not part of the program, our records indicate that they worked at the University of Pennsylvania, and we are curious how they did. Please answer the following question about overall performance for any youth you supervised or worked with this summer. **Please leave blank for any youth you did not supervise this summer.**

Overall, how would you rate the following youth as an employee?

	Very poor	Poor	Neutral	Good	Very good	Excellent	Exceptional
Patti Dennis	☐	☐	☐	☐	☐	☐	☐
Otis Elliott	☐	☐	☐	☐	☐	☐	☐
Juanita Guerrero	☐	☐	☐	☐	☐	☐	☐
Russell Higgins	☐	☐	☐	☐	☐	☐	☐
Ed White	☐	☐	☐	☐	☐	☐	☐

››

Figure A.4: Example Cover Letter to the Letter of Recommendation

Department of
Youth & Community
Development

November 1, 2017

Sara Heller
123 Fake Street
New York, NY 10003

Dear Sara,

This past summer you participated in a New York City summer program. This letter contains five copies of a letter of recommendation your supervisor wrote for you. [You should also have received a link to an electronic copy at [Student Email], in case you want to have an electronic version or print out more of copies of the letter.]

This year, some participants were included in a "letter of recommendation" program. You were included in this program, and your employer gave us feedback that could help you get a job or show your teachers your strengths. We hope you will show your letter of recommendation to your teachers, your guidance counselor, and potential employers (for example, by including it in job applications).

If you have any questions about the program, please see a description on our website here: https://www1.nyc.gov/assets/dycd/downloads/pdf/FAQs_Pilot_2017.pdf

If you have additional questions, you can contact our academic partners: Judd B. Kessler (judd.kessler@wharton.upenn.edu) at the University of Pennsylvania, and Sara Heller (hellersa@sas.upenn.edu).

Sincerely,

DYCD Team

The New York City Department of Youth and Community Development (DYCD) invests in a network of community-based organizations and programs to alleviate the effects of poverty and to provide opportunities for New Yorkers and communities to flourish.

Empowering Individuals • Strengthening Families • Investing in Communities

Notes: This cover letter accompanied five copies of the recommendation sent to youth. The text in brackets appeared when we had an email address on file for the youth.

Figure A.5: Example Job Advertisement Email

Penn
University of Pennsylvania

Youth Job Opening!

Dear **Sara**:

A professor at the University of Pennsylvania is looking for former NYC summer job program participants, like you, to **apply for a short-term and flexible job.**

This is an opportunity to **earn money and gain work experience** while helping improve future youth employment programs.

Those hired will **not** need to be on site at the University of Pennsylvania. All tasks and duties necessary to the position will be completed remotely (on an Internet capable computer or by mail).

Qualifications:

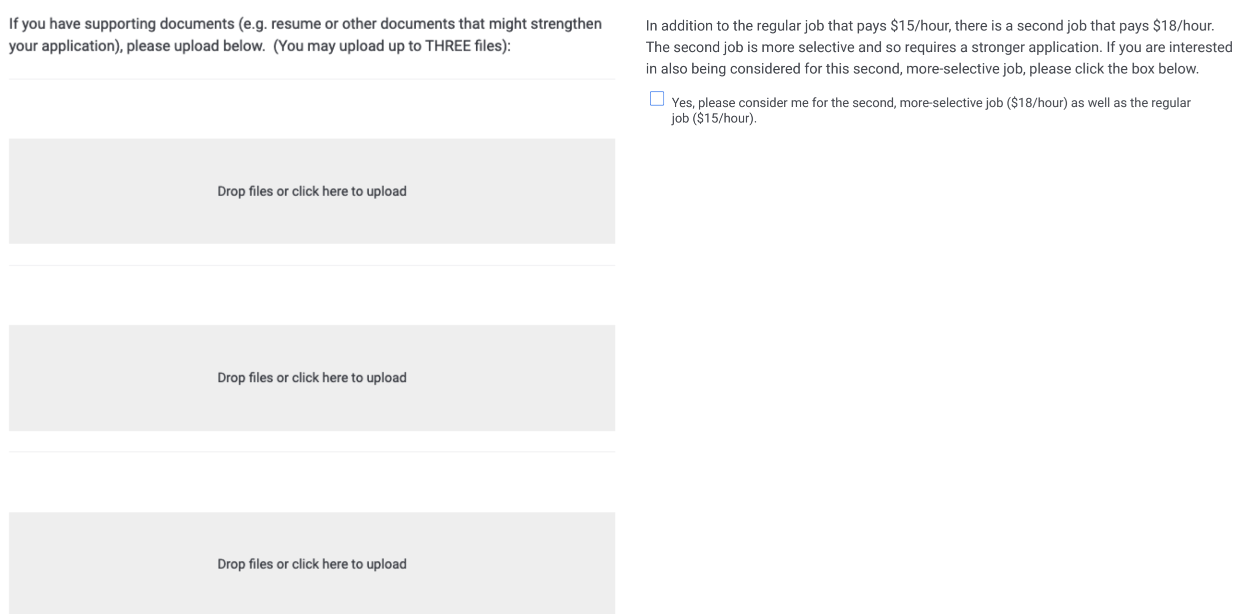
- Responsible
- Self-motivated
- Enthusiastic approach
- Some work experience preferred

Compensation for the job is $15/hour.

If you are **Sara Heller**, click this link to apply **(application due by March 30th)**:
Click here for your personal job application

All others who are interested can click this link for more information and to apply:
General job application

Figure A.6: Job Application Prompts to Upload Supporting Documents and to be Considered for More Selective Job



If you have supporting documents (e.g. resume or other documents that might strengthen your application), please upload below. (You may upload up to THREE files):

Drop files or click here to upload

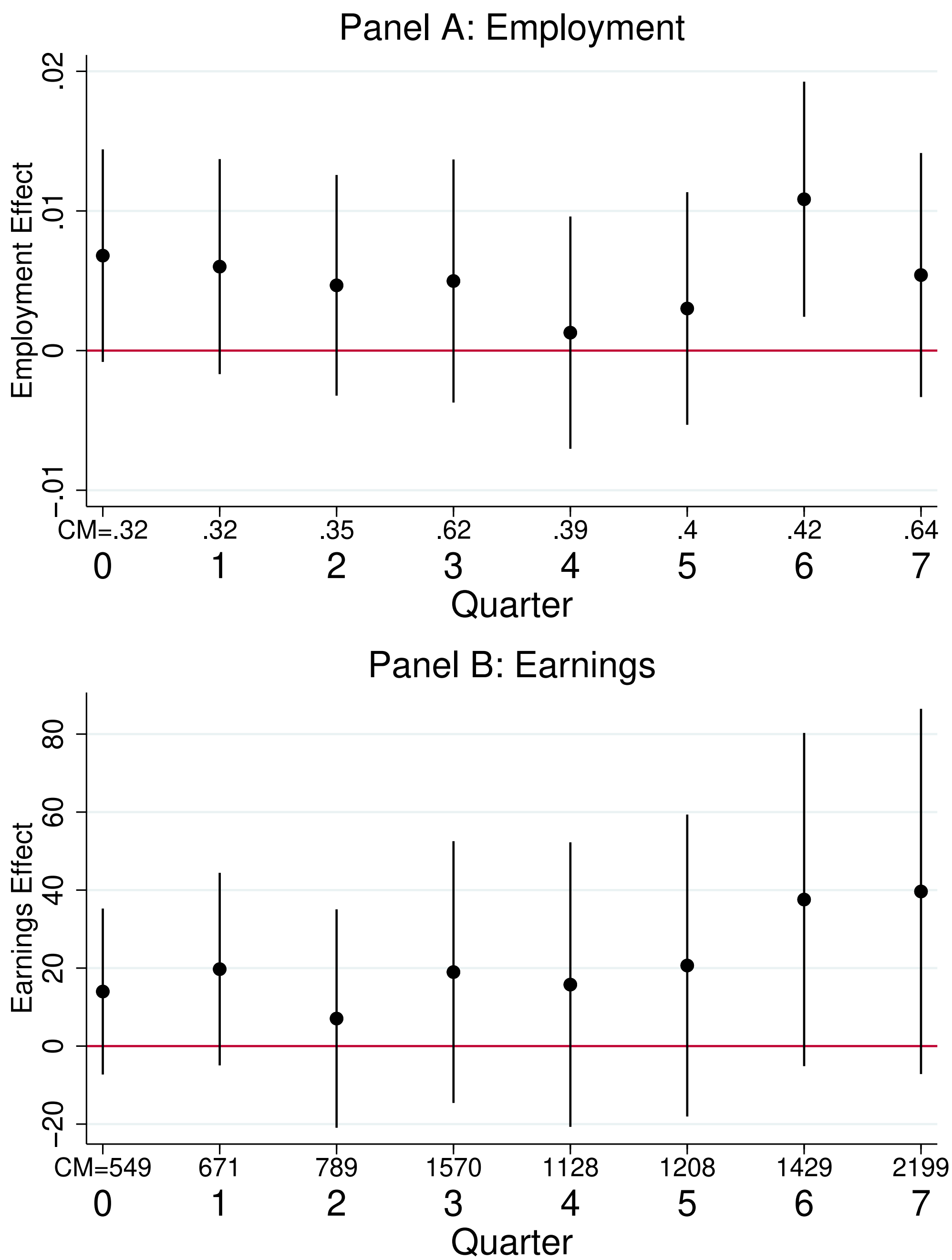
Drop files or click here to upload

Drop files or click here to upload

In addition to the regular job that pays $15/hour, there is a second job that pays $18/hour. The second job is more selective and so requires a stronger application. If you are interested in also being considered for this second, more-selective job, please click the box below.

Yes, please consider me for the second, more-selective job ($18/hour) as well as the regular job ($15/hour).

Figure A.7: Labor Market Effects by Quarter



Notes: Figure shows intent-to-treat effects on employment and earnings by quarter, with 95 percent confidence intervals calculated from cluster-robust standard errors. CM below axis displays the control mean in that quarter. All effects from regressions that include baseline covariates. Standard errors are clustered on individual. Quarters 0–3 comprise year 1 and quarters 4–7 comprise year 2.

Figure A.8: Available 4th- to 6th-Year Graduation Data Relative to Randomization, by Grade and Study Cohort

Pre-Randomization Grade	8th	9th	10th	11th	12th
Year Relative to Randomization			2016 Study Cohort		
-1 (graduated by 8/2016)					4th
1 (by 8/2017)				4th	5th
2 (by 8/2018)			4th	5th	6th
3 (by 8/2019)		4th	5th	6th	
4 (by 8/2020)	4th	5th	6th		
			2017 Study Cohort		
-1 (by 8/2017)					4th
1 (by 8/2018)				4th	5th
2 (by 8/2019)			4th	5th	6th
3 (by 8/2020)		4th	5th	6th	

= 10th–12th grade graduation sample

= 8th–12th grade sample with any graduation observed

Notes: Figure shows when 4th-, 5th-, and 6th-year graduation outcomes are observed for students in each pre-randomization grade level by study cohort. Gray boxes define our main 10th–12th grade graduation sample; black boxes are grades for which at least on-time graduation is observed within the "expected in high school" sample. Only 12th graders who had not graduated prior to letter distribution are included in these samples.

Table A.1: Earnings Impacts across Different Skewness Adjustments

Year	1	2	Cumulative
	Panel A: Winsorized at 99th Percentile		
ITT	57.80	109.62	166.88
	(44.86)	(70.87)	(103.63)
CM	3567	5913	9479
Sent Letter (IV)	148.61	280.14	426.40*
	(110.87)	(175.12)	(256.06)
CCM	3718	6112	9833
	Panel B: Log(Earnings + 0.1)		
ITT	0.124***	0.073	0.093**
	(0.042)	(0.044)	(0.037)
CM	4.92	5.44	6.97
Sent Letter (IV)	0.306***	0.18	0.230**
	(0.104)	(0.110)	(0.092)
CCM	4.94	5.56	7.02
	Panel C: Log(Earnings + 10)		
ITT	0.066***	0.045*	0.057**
	(0.024)	(0.026)	(0.022)
CM	6.30	6.73	7.7
Sent Letter (IV)	0.163***	0.112*	0.141**
	(0.058)	(0.064)	(0.055)
CCM	6.34	6.81	7.76
	Panel D: Log(Earnings + 100)		
ITT	0.038**	0.031*	0.039**
	(0.015)	(0.017)	(0.016)
CM	7.03	7.41	8.10
Sent Letter (IV)	0.094**	0.077*	0.095**
	(0.037)	(0.043)	(0.038)
CCM	7.07	7.47	8.15
	Panel E: Asinh(Earnings)		
ITT	0.104***	0.063*	0.080**
	(0.035)	(0.038)	(0.032)
CM	6.10	6.58	7.91
Sent Letter (IV)	0.256***	0.156*	0.198**
	(0.088)	(0.094)	(0.079)
CCM	6.12	6.69	7.97

Notes: N = 43,409. Winsorization in Panel A recodes each quarter's highest earnings to the 99th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.2: Spell Length and Censoring

	Total Spell Length	Spell Censored	Lasts at Least 2 Qtrs	Lasts at Least 3 Qtrs	Lasts at Least 4 Qtrs
			Spell 1		
ITT	0.0238	-0.0052	0.0010	0.0053	-0.0003
	(0.0289)	(0.0044)	(0.0048)	(0.0052)	(0.0051)
CM	3.22	0.23	0.63	0.47	0.37
IV	0.0592	-0.0129	0.0025	0.0134	-0.0002
	(0.0704)	(0.0107)	(0.0117)	(0.0125)	(0.0122)
CCM	3.41	0.25	0.65	0.50	0.41
N	36647	36647	33546	32626	31957
			Spell 2		
ITT	0.0435**	0.0011	0.0024	0.0135*	0.0180**
	(0.0187)	(0.0059)	(0.0072)	(0.0077)	(0.0074)
CM	1.99	0.49	0.63	0.38	0.26
IV	0.1058**	0.0023	0.0054	0.0324*	0.0425**
	(0.0453)	(0.0143)	(0.0171)	(0.0184)	(0.0177)
CCM	1.99	0.50	0.64	0.39	0.25
N	25203	25203	18003	15788	14154
			Spell 3		
ITT	-0.0043	0.0065	-0.0042	-0.0055	-0.0031
	(0.0208)	(0.0085)	(0.0105)	(0.0115)	(0.0110)
CM	1.77	0.55	0.61	0.35	0.22
IV	-0.0111	0.0154	-0.0124	-0.0132	-0.0073
	(0.0499)	(0.0203)	(0.0250)	(0.0276)	(0.0262)
CCM	1.81	0.56	0.64	0.39	0.23
N	12820	12820	8660	6903	5604

Notes: Total Spell Length conditions on youth having a spell. Indicators for at least X quarters are conditional on observing at least X quarters in the data. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.3: Labor Market Effects for DYCD and Non-DYCD Employers

	DYCD			Non-DYCD Employers		
Year	1	2	Cumulative	1	2	Cumulative
	Panel A: Employment					
ITT	0.0046	0.0033	0.0027	0.0088**	0.0008	0.0018
	(0.0046)	(0.0041)	(0.0046)	(0.0040)	(0.0043)	(0.0041)
CM	0.4157	0.262	0.5011	0.4256	0.5655	0.6218
Sent Letter (IV)	0.0119	0.0079	0.007	0.0221**	0.0021	0.0045
	(0.0114)	(0.0101)	(0.0114)	(0.0100)	(0.0106)	(0.0101)
CCM	0.419	0.253	0.507	0.429	0.588	0.639
	Panel B: Earnings, Winsorized at 99.5th Percentile					
ITT	0.92	3.04	4.11	61.05	105.82	166.86
	(10.68)	(9.94)	(16.66)	(46.23)	(73.89)	(108.01)
CM	810	572	1382	2770	5391	8161
Sent Letter (IV)	1.54	7.22	9.45	150.90	268.63	421.21
	(26.41)	(24.56)	(41.20)	(114.23)	(182.63)	(267.00)
CCM	870	574	1444	2861	5593	8453

N = 43,409. DYCD shows employment and earnings at employers with the FEIN of the agency that runs the SYEP. Non-DYCD shows all other employment. Winsorization in Panel B recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.4: Labor Market Effects by Industry Cluster

	SYEP-Related Industries (Cluster 1)			Other Industries (Cluster 2)		
Year	1	2	Cumulative	1	2	Cumulative
	Panel A: Employment					
ITT	0.0049	0.0063	0.0026	0.0071*	0.001	0.0007
	(0.0047)	(0.0047)	(0.0045)	(0.0040)	(0.0044)	(0.0044)
CM	0.5243	0.4407	0.6495	0.3105	0.4256	0.4869
Sent Letter (IV)	0.0119	0.015	0.0061	0.0181*	0.003	0.0022
	(0.0116)	(0.0116)	(0.0111)	(0.0099)	(0.0110)	(0.0108)
CCM	0.536	0.443	0.665	0.307	0.436	0.496
	Panel B: Earnings, Winsorized at 99.5th Percentile					
ITT	19.01	3.09	21.53	34.22	108.76*	135.06
	(31.19)	(49.22)	(71.96)	(40.93)	(65.37)	(95.57)
CM	1658	2261	3919	1882	3637	5518
Sent Letter (IV)	51.02	8.13	60.24	85.19	275.02*	341.83
	(77.07)	(121.69)	(177.89)	(101.20)	(161.58)	(236.30)
CCM	1812	2451	4262	1881	3621	5521

N = 43,409. Industry definition follows the cluster definitions in Gelber, Isen, and Kessler 2016. SYEP-related include employment in industries that are over-represented among summer jobs in the program. Other industries are those under-represented in summer jobs. Winsorization in Panel B recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.5: Descriptive Statistics, Job Application Sample

	Control Mean	Treatment Mean	Test of Difference
N	2000	2000	
Age	17.1	17.2	0.205
Male	0.421	0.444	0.151
Black	0.405	0.385	0.196
Hispanic	0.290	0.328	0.010
Asian	0.140	0.125	0.160
White	0.125	0.124	0.928
Other Race	0.040	0.038	0.789
In High School	0.759	0.753	0.686
HS Graduate	0.045	0.043	0.758
In College	0.173	0.173	0.967
Not in UI Data	0.005	0.013	0.011
Never Employed Pre-SYEP	0.448	0.456	0.611
Ever Worked, Year -4	0.149	0.166	0.129
Earnings, Year -4	294	343	0.425
Ever Worked, Year -3	0.278	0.279	0.944
Earnings, Year -3	605	626	0.777
Ever Worked, Year -2	0.439	0.445	0.726
Earnings, Year -2	1122	1093	0.759
Ever Worked, Year -1	0.976	0.968	0.105
Earnings, Year -1	2519	2496	0.842
No Education Match	0.136	0.136	0.963
In HS Sample	0.451	0.449	0.899
Joint F-test	$F(24, 3975) = 1.216$, $p=.215$		

N = 4,000. Test of difference reports the p-value from a regression of each characteristic on a treatment indicator, using standard errors clustered on individual.

Table A.6: Balance for All Rated Youth by Rating Group

	Control Low	Treatment Low	Test of Difference	Control High	Treatment High	Test of Difference
N	5062	4632		10425	9768	
Age	17.14	17.06	0.084	17.25	17.25	1.000
Male	0.449	0.448	0.935	0.414	0.417	0.753
Black	0.492	0.500	0.419	0.382	0.371	0.118
Hispanic	0.292	0.294	0.836	0.284	0.287	0.678
Asian	0.099	0.091	0.224	0.147	0.159	0.015
White	0.069	0.070	0.875	0.140	0.137	0.566
Other Race	0.049	0.045	0.392	0.047	0.045	0.598
In High School	0.782	0.787	0.549	0.739	0.734	0.371
HS Graduate	0.046	0.043	0.469	0.040	0.040	0.768
In College	0.133	0.132	0.950	0.204	0.208	0.390
Not in UI Data	0.009	0.009	0.902	0.003	0.004	0.696
Never Employed Pre-SYEP	0.461	0.489	0.007	0.438	0.437	0.871
Ever Worked, Year -4	0.145	0.129	0.024	0.159	0.159	0.964
Earnings, Year -4	272	269	0.935	343	353	0.728
Ever Worked, Year -3	0.255	0.236	0.037	0.274	0.282	0.177
Earnings, Year -3	507	477	0.432	618	642	0.487
Ever Worked, Year -2	0.424	0.403	0.037	0.450	0.454	0.523
Earnings, Year -2	995	879	0.024	1117	1153	0.397
Ever Worked, Year -1	0.965	0.979	0.000	0.989	0.993	0.014
Earnings, Year -1	2200	2134	0.269	2519	2583	0.216
No Education Match	0.094	0.089	0.399	0.131	0.130	0.846
In HS Sample	0.488	0.494	0.542	0.440	0.441	0.826
Joint F-test	F(24, 9587) = 1.397, p=.094			F(24, 19643) = .725, p=.831		

Notes: Sample includes all youth with employer rating (N = 29,887, 256 youth missing race/ethnicity). Low includes rating categories 1–4; High includes rating categories 5–7. Test of difference reports the p-value from a regression of each characteristic on a treatment indicator within that rating group, controlling for a cohort indicator and using standard errors clustered on individual.

Table A.7: Labor Market Effects for Youth with High and Low Employer Ratings, All Rated Youth

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Low Ratings		0.0136	0.0006	0.0075	12.68	26.53	34.89
		(0.0089)	(0.0089)	(0.0073)	(88.92)	(143.15)	(208.89)
ITT, High Ratings		0.0123**	0.0099*	0.0101**	138.15*	187.37*	324.33**
		(0.0059)	(0.0060)	(0.0047)	(70.52)	(112.15)	(164.20)
P-value, test of diff.		0.905	0.383	0.765	0.269	0.376	0.276
CM, Low		0.673	0.721	0.836	3109	5409	8518
CM, High		0.715	0.720	0.846	3729	6251	9979
	First Stage						
IV, Low Ratings	0.3067***	0.0438	0.0017	0.024	22.96	92.65	114.00
	(0.0068)	(0.0290)	(0.0290)	(0.0238)	(289.28)	(466.86)	(681.24)
IV, High Ratings	0.7529***	0.0162**	0.0131*	0.0134**	181.27*	251.10*	431.32**
	(0.0043)	(0.0079)	(0.0079)	(0.0063)	(93.59)	(148.98)	(218.01)
P-value, test of diff.	0.000	0.359	0.704	0.666	0.602	0.746	0.657
CCM, Low		0.642	0.71	0.802	3180	5343	8525
CCM, High		0.713	0.733	0.851	3576	6241	9818

Notes: Sample includes all youth who were rated on a survey (N = 29,887). Low Ratings includes rating categories 1–4; High Ratings includes rating categories 5–7. Test of difference shows p-value for null hypothesis that treatment effects are equal in low-rated and high-rated groups. Winsorization in Panel B recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.8: Balance by Rating Group, Fully Completed Surveys

	Control Low	Treatment Low	Test of Difference	Control High	Treatment High	Test of Difference
N	2209	2092		4833	4777	
Age	17.09	17.09	0.919	17.26	17.23	0.453
Male	0.440	0.439	0.937	0.400	0.409	0.352
Black	0.505	0.535	0.053	0.388	0.381	0.481
Hispanic	0.277	0.258	0.178	0.286	0.292	0.491
Asian	0.117	0.111	0.573	0.165	0.178	0.090
White	0.051	0.047	0.465	0.114	0.105	0.145
Other Race	0.050	0.049	0.874	0.047	0.044	0.461
In High School	0.785	0.783	0.846	0.735	0.736	0.941
HS Graduate	0.042	0.041	0.808	0.037	0.031	0.141
In College	0.137	0.139	0.890	0.211	0.216	0.550
Not in UI Data	0.007	0.008	0.877	0.004	0.004	0.597
Never Employed Pre-SYEP	0.481	0.481	0.969	0.450	0.462	0.222
Ever Worked, Year -4	0.134	0.125	0.392	0.150	0.149	0.963
Earnings, Year -4	268	266	0.974	346	320	0.527
Ever Worked, Year -3	0.242	0.234	0.540	0.259	0.266	0.394
Earnings, Year -3	466	479	0.838	603	594	0.854
Ever Worked, Year -2	0.407	0.395	0.435	0.435	0.426	0.388
Earnings, Year -2	933	839	0.224	1050	1060	0.851
Ever Worked, Year -1	0.971	0.980	0.048	0.989	0.991	0.327
Earnings, Year -1	2102	2063	0.650	2475	2410	0.342
No Education Match	0.083	0.078	0.552	0.111	0.114	0.621
In HS Sample	0.498	0.493	0.736	0.460	0.455	0.660
Joint F-test	$F(24, 4264) = .897$, $p=.607$			$F(24, 9471) = .91$, $p=.588$		

Notes: Sample includes all youth on a fully completed survey (N=13,911, 167 youth missing race/ethnicity). Low includes rating categories 1–4; High includes rating categories 5–7. Test of difference reports the p-value from a regression of each characteristic on a treatment indicator within that rating group, controlling for a cohort indicator and using standard errors clustered on individual.

Table A.9: Education Descriptive Statistics

	N	Control Mean	Treatment Mean	Test of Difference	N	Control Mean	Treatment Mean	Test of Difference
	Expected in HS Sample				Main Graduation Sample (10-12th)			
Age	19714	15.96	15.95	0.357	13732	16.46	16.44	0.308
Male	19714	0.452	0.445	0.344	13732	0.442	0.433	0.294
Black	19656	0.426	0.424	0.854	13674	0.408	0.410	0.784
Hispanic	19656	0.309	0.307	0.821	13674	0.314	0.310	0.756
Asian	19656	0.139	0.137	0.794	13674	0.151	0.148	0.621
White	19656	0.074	0.084	0.009	13674	0.074	0.084	0.032
Grade Level	19714	10.04	10.03	0.344	13732	10.60	10.59	0.166
Share Enrolled Days Present	19714	0.902	0.899	0.169	13732	0.906	0.901	0.046
Missing GPA	19714	0.100	0.101	0.848	13732	0.035	0.034	0.746
GPA (100 point scale)	17732	79.73	79.34	0.033	13259	80.55	80.05	0.014
In Graduation Sample (8-12th)	19714	0.942	0.939	0.316	13732	1	1	--
In College Sample	19714	0.711	0.709	0.756	13732	0.945	0.949	0.204
Not in UI Data	19714	0.013	0.015	0.275	13732	0.008	0.009	0.506
Never Employed Pre-SYEP	19714	0.614	0.621	0.271	13732	0.524	0.535	0.233
Ever Worked, Year -4	19714	0.041	0.040	0.688	13732	0.055	0.053	0.617
Earnings, Year -4	19714	77	96	0.342	13732	104	98	0.816
Ever Worked, Year -3	19714	0.134	0.134	0.933	13732	0.177	0.178	0.944
Earnings, Year -3	19714	180	193	0.555	13732	243	227	0.562
Ever Worked, Year -2	19714	0.305	0.304	0.865	13732	0.373	0.371	0.793
Earnings, Year -2	19714	431	421	0.682	13732	547	499	0.118
Ever Worked, Year -1	19714	0.959	0.960	0.614	13732	0.963	0.968	0.140
Earnings, Year -1	19714	1579	1563	0.594	13732	1763	1703	0.109
Joint F-test	F(37, 19063) = 1.239, p=.151				F(35, 13378) = 1.366, p=.073			

Notes: Left panel shows statistics for the expected in high school sample; right panel shows statistics for the graduation sample (see text for details). Test of difference reports the p-value from a regression of each characteristic on a treatment indicator, controlling for a cohort indicator and using standard errors clustered on individual.

Table A.10: Employment and Earnings Effects for the Expected in HS Sample

	First Stage	Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Expected in HS Data		0.0144**	0.0126**	0.0098*	9.74	77.99	80.75
		(0.0065)	(0.0064)	(0.0054)	(43.58)	(77.30)	(107.19)
ITT, Not Expected in HS Data		0.0112**	0.0001	0.0065	104.26	136.85	241.59
		(0.0052)	(0.0053)	(0.0042)	(75.95)	(117.94)	(174.80)
P-value, test of diff.		0.696	0.132	0.624	0.281	0.677	0.433
CM, Exp. In HS		0.636	0.677	0.810	2120	3908	6028
CM, Not Exp. In HS		0.755	0.756	0.866	4794	7675	12469
IV, Expected in HS Data	0.4138***	0.0349**	0.0306**	0.0237*	31.73	208.38	239.09
	(0.0049)	(0.0158)	(0.0155)	(0.0130)	(105.45)	(186.83)	(259.02)
IV, Not Expected in HS Data	0.3966***	0.0282**	0.0002	0.0157	259.84	345.29	604.01
	(0.0045)	(0.0132)	(0.0135)	(0.0106)	(191.00)	(297.03)	(440.06)
P-value, test of diff.	0.01	0.743	0.139	0.632	0.296	0.697	0.476
CCM, Exp. in HS		0.644	0.680	0.819	2302	4072	6375
CCM, Not Exp. In HS		0.743	0.770	0.860	4960	7964	12924

Notes: Analysis is conducted on all those expected to be observed based on pre-program grade of enrollment (see text for details). Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.11: Education Effects

	Ever Enrolled Y1	Days Enrolled Y1	Share Days Present Y1	GPA Y1	On-time Graduation	Ever Graduated	Graduated or Still Attending	On-time College
	Panel A: All Expected in High School (8th - 12th Graders)							
ITT	-0.0019	-0.3152	0.0014	-0.1300	-0.0064	-0.0023	-0.0031	0.0005
	(0.0030)	(0.5444)	(0.0028)	(0.0988)	(0.0041)	(0.0040)	(0.0037)	(0.0062)
CM	0.946	150.0	0.829	80.13	0.791	0.836	0.887	0.697
Sent Letter (IV)	-0.0045	-0.7681	0.0033	-0.3025	-0.015	-0.0051	-0.0073	0.0014
	(0.0072)	(1.3126)	(0.0067)	(0.2367)	(0.0098)	(0.0094)	(0.0089)	(0.0141)
CCM	0.957	153.1	0.849	81.75	0.833	0.869	0.919	0.721
N	19714	19714	19714	18237	18537	18537	18537	13999
	Panel B: Main Graduation Sample (10th - 12th Graders)							
ITT	-0.0053	-0.9527	-0.0025	-0.1623	-0.0087*	-0.0011	-0.002	-0.0002
	(0.0035)	(0.6467)	(0.0032)	(0.1100)	(0.0044)	(0.0042)	(0.0040)	(0.0064)
CM	0.946	150.1	0.8317	81.397	0.815	0.867	0.900	0.705
Sent Letter (IV)	-0.0116	-2.0855	-0.0056	-0.3572	-0.0194*	-0.0022	-0.0045	-0.0002
	(0.0080)	(1.4616)	(0.0073)	(0.2473)	(0.0100)	(0.0094)	(0.0091)	(0.0145)
CCM	0.963	154	0.855	82.62	0.848	0.886	0.922	0.726
N	13732	13732	13732	12677	13732	13732	13732	13005

Notes: Analysis is conducted on all those expected to be observed based on pre-program grade of enrollment (see text for details). On-time graduation equals 1 for public school, non-charter students who graduate within 4 years and do not transfer to GED programs or other districts. Ever graduated includes any 5th- and 6th-year graduation observed during the follow-up period. 8th–12th graders are included in both measures. College enrollment is only measured within 6 months after a student's on-time graduation date, regardless of graduation status. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.12: Education Effects for Minority and White Youth

	First Stage	Ever Enrolled Y1	Days Enrolled Y1	Share Days Present Y1	GPA Y1	On-time Graduation	Ever Graduated	Graduated or Still Attending	On-time College
ITT, Minority		-0.0017	-0.2870	0.0017	-0.1389	-0.0076*	-0.0032	-0.0038	-0.0004
		(0.0032)	(0.5791)	(0.0029)	(0.1041)	(0.0044)	(0.0042)	(0.0040)	(0.0065)
ITT, White		-0.0017	-0.5683	0.0021	0.0172	0.0082	0.0082	0.0031	0.0076
		(0.0082)	(1.4672)	(0.0079)	(0.3317)	(0.0107)	(0.0103)	(0.0101)	(0.0190)
P-value, test of diff.		0.997	0.859	0.963	0.653	0.171	0.304	0.519	0.692
CM, Minority		0.943	149.3	0.822	79.50	0.783	0.830	0.882	0.686
CM, White		0.974	158.6	0.909	87.70	0.888	0.902	0.948	0.824
IV, Minority	0.4193***	-0.004	-0.6575	0.004	-0.3361	-0.0177*	-0.0073	-0.0088	-0.0009
	(0.0052)	(0.0076)	(1.3777)	(0.0070)	(0.2456)	(0.0102)	(0.0099)	(0.0093)	(0.0147)
IV, White	0.3407***	-0.0046	-1.5909	0.0062	0.0202	0.0257	0.0256	0.0100	0.0255
	(0.0166)	(0.0242)	(4.3503)	(0.0234)	(0.9784)	(0.0319)	(0.0306)	(0.0298)	(0.0545)
P-value, test of diff.	0.000	0.980	0.838	0.930	0.724	0.195	0.308	0.548	0.641
CCM, Minority		0.955	152.5	0.843	81.18	0.828	0.867	0.917	0.713
CCM, White		0.987	160.4	0.907	88.99	0.884	0.895	0.945	0.806

Notes: Analysis is conducted on all those expected to be observed based on pre-program grade of enrollment (see text for details). On-time graduation equals 1 for public school, non-charter students who graduate within 4 years and do not transfer to GED programs or other districts. Ever graduated includes any 5th- and 6th-year graduation observed during the follow-up period. 8th–12th graders are included in both measures. College enrollment is only measured within 6 months after a student's on-time graduation date, regardless of graduation status. For those in the expected in high school sample, there are 18,238 minority youth and 1,553 White youth, with 58 observations dropped due to missing race/ethnicity. For the graduation sample, there are 16,986 minority youth and 1,493 White youth, with 58 observations dropped due to missing race/ethnicity. For college sample, there are 12,852 minority youth and 1,158 White youth, with 56 observations dropped due to missing race/ethnicity. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.13: Effects on Joint Employment and Graduation Outcomes, Grades 10-12, ITT

	Panel A: On-Time Graduation			
	Ever Work, On-time Grad	Never Work, On-time Grad	Ever Work, Not On-time	Never Work, Not On-time
ITT	-0.0001	-0.0082	0.0122***	-0.0042
	(0.0064)	(0.0052)	(0.0045)	(0.0031)
CM	0.703	0.112	0.144	0.041
	Panel B: Any Graduation			
	Ever Work, Graduated	Never Work, Graduated	Ever Work, Not Graduated	Never Work, Not Graduated
ITT	0.0095	-0.0105**	0.0027	-0.0017
	(0.0064)	(0.0053)	(0.0040)	(0.0029)
CM	0.748	0.120	0.099	0.033
	Panel C: Any Graduation or Continued Attendance			
	Ever Work, Persisted	Never Work, Persisted	Ever Work, Not Persisted	Never Work, Not Persisted
ITT	0.0058	-0.0078	0.0064*	-0.0045*
	(0.0064)	(0.0054)	(0.0037)	(0.0026)
CM	0.775	0.125	0.072	0.028

Notes: N=13,732. Analysis is conducted on main graduation sample (non-charter 10th–12th graders in the pre-randomization year, see text for details). The first stage for this subsample is 0.44. Panel A shows whether someone ever worked during the two-year follow up and whether they graduated on-time during the data. Panel B shows whether someone ever worked during the two-year follow up and whether we ever observed them graduate in the data, regardless of timing. Panel C shows whether someone ever worked during the two-year follow up and whether they either graduated in the data or had positive days attended in the last year in the data. CM shows control means. Coefficients may not total to 0 due to the inclusion of different sets of covariates in the post double-selection LASSO. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.14: Joint Outcomes for Minority and White Youth, Grades 10-12, ITT

	Panel A: On-Time Graduation			
	Ever Work, On-time	Never Work, On-time	Ever Work, Not On-time	Never Work, Not On-time
ITT, Minority	0.0014	-0.0108**	0.0145***	-0.0055
	(0.0067)	(0.0053)	(0.0049)	(0.0033)
ITT, White	-0.0149	0.0163	-0.0144	0.011
	(0.0223)	(0.0205)	(0.0118)	(0.0090)
P-value, test of diff.	0.485	0.200	0.024	0.087
CM, Minority	0.698	0.111	0.149	0.042
CM, White	0.758	0.134	0.087	0.022
	Panel B: Any Graduation			
	Ever Work, Graduated	Never Work, Graduated	Ever Work, Not Graduated	Never Work, Not Graduated
ITT, Minority	0.0115*	-0.0137**	0.0048	-0.0025
	(0.0067)	(0.0055)	(0.0043)	(0.0031)
ITT, White	-0.0127	0.0196	-0.0181*	0.0076
	(0.0222)	(0.0207)	(0.0101)	(0.0084)
P-value, test of diff.	0.297	0.120	0.038	0.26
CM, Minority	0.745	0.118	0.102	0.035
CM, White	0.774	0.136	0.071	0.020
	Panel C: Any Graduation or Continued Attendance			
	Ever Work, Persisted	Never Work, Persisted	Ever Work, Not Persisted	Never Work, Not Persisted
ITT, Minority	0.0077	-0.0104*	0.0086**	-0.0057**
	(0.0067)	(0.0056)	(0.0040)	(0.0028)
ITT, White	-0.0142	0.018	-0.015	0.0088
	(0.0228)	(0.0212)	(0.0102)	(0.0077)
P-value, test of diff.	0.358	0.195	0.032	0.075
CM, Minority	0.773	0.123	0.075	0.030
CM, White	0.805	0.144	0.039	0.012

Notes: N = 12,589 Minority youth and N = 1,085 White youth, with 58 observations in graduation data dropped due to missing race/ethnicity. Sample, outcomes, and first stage reported in Table 8. CM shows control means. Coefficients may not total to 0 due to the inclusion of different sets of covariates in the post double-selection LASSO. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.15: Effects on Joint Outcomes by GPA, Grades 10-12, ITT

	Panel A: On-Time Graduation			
	Ever Work, On-time	Never Work, On-time	Ever Work, Not On-time	Never Work, Not On-time
ITT, Below Median GPA	-0.0095 (0.0106)	-0.0188*** (0.0067)	0.0320*** (0.0092)	-0.0054 (0.0056)
ITT, Above Median GPA	-0.001 (0.0090)	0.0004 (0.0084)	-0.0002 (0.0033)	0.0012 (0.0021)
P-value, test of diff.	0.544	0.072	0.001	0.270
CM, Below	0.621	0.092	0.228	0.059
CM, Above	0.833	0.140	0.020	0.007
	Panel B: Any Graduation			
	Ever Work, Graduated	Never Work, Graduated	Ever Work, Not Graduated	Never Work, Not Graduated
ITT, Below Median GPA	0.0132 (0.0103)	-0.0224*** (0.0071)	0.0096 (0.0079)	-0.0017 (0.0050)
ITT, Above Median GPA	-0.0007 (0.0088)	0.0000 (0.0084)	-0.0009 (0.0027)	0.0016 (0.0021)
P-value, test of diff.	0.303	0.041	0.209	0.531
CM, Below	0.707	0.106	0.143	0.044
CM, Above	0.840	0.141	0.013	0.007
	Panel C: Any Graduation or Continued Attendance			
	Ever Work, Persisted	Never Work, Persisted	Ever Work, Not Persisted	Never Work, Not Persisted
ITT, Below Median GPA	0.0074 (0.0101)	-0.0183** (0.0074)	0.0158** (0.0073)	-0.0055 (0.0045)
ITT, Above Median GPA	0.0004 (0.0087)	0.0009 (0.0084)	-0.0016 (0.0023)	0.0006 (0.0018)
P-value, test of diff.	0.601	0.085	0.023	0.206
CM, Below	0.737	0.113	0.113	0.038
CM, Above	0.843	0.142	0.010	0.005

Notes: N = 13,259, 473 youth dropped for missing GPA. Median GPA in this sample is 81.64. Analysis on main graduation sample (non-charter 10th–12th graders in the pre-randomization year, see text for details). First stage reported in the following table. Panel A shows whether someone ever worked during the two-year follow up and whether they graduated on-time during the data. Panel B shows whether someone ever worked during the two-year follow up and whether we ever observed them graduate in the data, regardless of timing. Panel C shows whether someone ever worked during the two-year follow up and whether they either graduated in the data or had positive days attended in the last year in the data. CM shows control means. Coefficients may not total to 0 due to the inclusion of different sets of covariates in the post double-selection LASSO. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.16: Effects on Joint Outcomes by GPA, Grades 10-12, IV

		Panel A: On-Time Graduation			
		Ever Work, On-time	Never Work, On-time	Ever Work, Not On-time	Never Work, Not On-time
	First Stage				
IV, Below Median	0.3945***	-0.0233	-0.0490***	0.0819***	-0.0139
GPA	(0.0084)	(0.0267)	(0.0169)	(0.0234)	(0.0140)
IV, Above Median	0.4804***	0.0011	-0.0017	-0.0008	0.0013
GPA	(0.0087)	(0.0186)	(0.0173)	(0.0068)	(0.0045)
P-value, test of diff.	0.00	0.454	0.051	0.001	0.304
CCM, Below		0.675	0.121	0.160	0.049
CCM, Above		0.849	0.131	0.014	0.006
		Panel B: Any Graduation			
		Ever Work, Graduated	Never Work, Graduated	Ever Work, Not Graduated	Never Work, Not Graduated
IV, Below Median		0.0343	-0.0581***	0.0251	-0.0044
GPA		(0.0259)	(0.0179)	(0.0201)	(0.0125)
IV, Above Median		0.002	-0.0028	-0.0021	0.0026
GPA		(0.0183)	(0.0174)	(0.0057)	(0.0043)
P-value, test of diff.		0.31	0.027	0.195	0.597
CCM, Below		0.732	0.135	0.103	0.034
CCM, Above		0.851	0.132	0.012	0.004
		Panel C: Any Graduation or Continued Attendance			
		Ever Work, Persisted	Never Work, Persisted	Ever Work, Not Persisted	Never Work, Not Persisted
IV, Below Median		0.0196	-0.0479***	0.0402**	-0.0141
GPA		(0.0254)	(0.0186)	(0.0185)	(0.0114)
IV, Above Median		0.0044	-0.0011	-0.0036	0.0006
GPA		(0.0181)	(0.0174)	(0.0048)	(0.0038)
P-value, test of diff.		0.626	0.067	0.022	0.221
CCM, Below		0.77	0.134	0.064	0.035
CCM, Above		0.853	0.133	0.009	0.004

Notes: N=13,259, 473 youth dropped for missing GPA. Median GPA in this sample is 81.64. Analysis on main graduation sample (non-charter 10th–12th graders in the pre-randomization year, see text for details). Panel A shows whether someone ever worked during the two-year follow up and whether they graduated on-time during the data. Panel B shows whether someone ever worked during the two-year follow up and whether we ever observed them graduate in the data, regardless of timing. Panel C shows whether someone ever worked during the two-year follow up and whether they either graduated in the data or had positive days attended in the last year in the data. CCM shows control complier means, which may not total to 1 across categories due to estimation error in the IV and the inclusion of different sets of covariates in the post double-selection LASSO. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.17: Effects on Joint Employment and Graduation Outcomes, Grades 8-12, ITT

	Panel A: On-Time Graduation			
	Ever Work, On-time Grad	Never Work, On-time Grad	Ever Work, Not On-time	Never Work, Not On-time
ITT	-0.0010	-0.0054	0.0069*	-0.0011
	(0.0058)	(0.0047)	(0.0041)	(0.0031)
CM	0.667	0.125	0.156	0.052
	Panel B: Any Graduation			
	Ever Work, Graduated	Never Work, Graduated	Ever Work, Not Graduated	Never Work, Not Graduated
ITT	0.0047	-0.0067	0.0016	0.0001
	(0.0058)	(0.0048)	(0.0038)	(0.0030)
CM	0.705	0.131	0.118	0.046
	Panel C: Any Graduation or Continued Attendance			
	Ever Work, Persisted	Never Work, Persisted	Ever Work, Not Persisted	Never Work, Not Persisted
ITT	0.0000	-0.0032	0.0060*	-0.0029
	(0.0058)	(0.0050)	(0.0034)	(0.0025)
CM	0.744	0.143	0.079	0.034

Notes: N=18,537. Analysis on expanded graduation sample (non-charter 8th–12th graders in the pre-randomization year, see text for details). First stage for this subsample reported in following table. Panel A shows whether someone ever worked during the two-year follow up and whether they graduated on-time during the data. Panel B shows whether someone ever worked during the two-year follow up and whether we ever observed them graduate in the data, regardless of timing. Panel C shows whether someone ever worked during the two-year follow up and whether they either graduated in the data or had positive days attended in the last year in the data. CM shows control means. Coefficients may not total to 0 due to the inclusion of different sets of covariates in the post double-selection LASSO. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.18: Effects on Joint Employment and Graduation Outcomes, Grades 8-12, IV

		Panel A: On-Time Graduation			
		Ever Work, On-time Grad	Never Work, On-time Grad	Ever Work, Not On-time	Never Work, Not On-time
	First Stage				
Sent Letter (IV)	0.4189***	-0.0026	-0.0133	0.0170*	-0.0023
	(0.0051)	(0.0137)	(0.0111)	(0.0098)	(0.0074)
CCM		0.710	0.124	0.130	0.038
		Panel B: Any Graduation			
		Ever Work, Graduated	Never Work, Graduated	Ever Work, Not Graduated	Never Work, Not Graduated
Sent Letter (IV)		0.0114	-0.0164	0.0045	0.0005
		(0.0138)	(0.0114)	(0.0091)	(0.0071)
CCM		0.739	0.130	0.099	0.032
		Panel C: Any Graduation or Continued Attendance			
		Ever Work, Persisted	Never Work, Persisted	Ever Work, Not Persisted	Never Work, Not Persisted
Sent Letter (IV)		0.0003	-0.0078	0.0147*	-0.0075
		(0.0139)	(0.0119)	(0.0081)	(0.0060)
CCM		0.784	0.135	0.055	0.026

Notes: N=18,537. Analysis on expanded graduation sample (non-charter 8th–12th graders in the pre-randomization year, see text for details). First stage for this subsample reported in following table. Panel A shows whether someone ever worked during the two-year follow up and whether they graduated on-time during the data. Panel B shows whether someone ever worked during the two-year follow up and whether we ever observed them graduate in the data, regardless of timing. Panel C shows whether someone ever worked during the two-year follow up and whether they either graduated in the data or had positive days attended in the last year in the data. CCM shows control complier means, which may not total to 1 across categories due to estimation error in the IV and the inclusion of different sets of covariates in the post double-selection LASSO. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.19: Earnings Impacts Across Different Skewness Adjustments by Race/Ethnicity

Year	1	2	Cumulative	1	2	Cumulative
	White			Minority		
	Panel A: Winsorized at 99th Percentile					
ITT	-66.60	-115.62	-180.29	76.39	142.25*	218.02**
	(139.80)	(208.70)	(314.35)	(47.28)	(75.37)	(109.75)
CM	3725	5589	9314	3530	5920	9450
Sent Letter (IV)	-230.03	-407.23	-631.33	187.92*	348.55*	534.18**
	(471.52)	(703.61)	(1059.81)	(112.82)	(179.85)	(261.86)
CCM	4353	6402	10750	3638	6051	9692
	Panel B: Log(Earnings + 0.1)					
ITT	0.039	-0.017	-0.032	0.131***	0.084*	0.106***
	(0.115)	(0.131)	(0.105)	(0.045)	(0.048)	(0.040)
CM	5.37	5.10	6.99	4.85	5.48	6.95
Sent Letter (IV)	0.127	-0.077	-0.117	0.313***	0.200*	0.252***
	(0.388)	(0.442)	(0.354)	(0.108)	(0.114)	(0.095)
CCM	5.51	5.38	7.26	4.88	5.57	7.00
	Panel C: Log(Earnings + 10)					
ITT	0.017	-0.009	-0.017	0.070***	0.052*	0.065***
	(0.064)	(0.076)	(0.063)	(0.025)	(0.028)	(0.024)
CM	6.52	6.51	7.68	6.27	6.76	7.70
Sent Letter (IV)	0.055	-0.042	-0.063	0.167***	0.125*	0.154***
	(0.217)	(0.257)	(0.213)	(0.061)	(0.067)	(0.058)
CCM	6.65	6.70	7.88	6.31	6.82	7.75
	Panel D: Log(Earnings + 100)					
ITT	0.007	-0.005	-0.007	0.041**	0.036*	0.043***
	(0.040)	(0.050)	(0.044)	(0.016)	(0.019)	(0.017)
CM	7.13	7.25	8.05	7.01	7.42	8.10
Sent Letter (IV)	0.022	-0.026	-0.029	0.097**	0.086*	0.103***
	(0.136)	(0.169)	(0.147)	(0.038)	(0.044)	(0.040)
CCM	7.25	7.39	8.20	7.05	7.47	8.14
	Panel E: Asinh(Earnings)					
ITT	0.031	-0.014	-0.027	0.110***	0.072*	0.091***
	(0.097)	(0.112)	(0.090)	(0.038)	(0.041)	(0.034)
CM	6.47	6.29	7.92	6.04	6.61	7.91
Sent Letter (IV)	0.102	-0.064	-0.099	0.262***	0.173*	0.218***
	(0.328)	(0.377)	(0.304)	(0.091)	(0.097)	(0.082)
CCM	6.60	6.54	8.17	6.07	6.70	7.96

Notes: N = 43,019, 390 youth excluded for missing race. Winsorization in Panel A recodes each quarter's highest earnings to the 99th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.20: Amount and Timing of Work by Race/Ethnicity

	Num Quarters Worked	Num of Job Spells	Num of Job Spells if >0	Time to First Qtr Worked
ITT, Minority	0.050**	0.021	0.001	-0.058***
	(0.023)	(0.015)	(0.015)	(0.022)
ITT, White	-0.009	-0.007	0.000	0.015
	(0.059)	(0.036)	(0.035)	(0.049)
P-value, test of diff.	0.354	0.458	0.971	0.175
CM, Minority	3.49	1.99	2.38	2.18
CM, White	3.17	1.90	2.23	2.28
IV, Minority	0.120**	0.051	0.005	-0.138***
	(0.056)	(0.036)	(0.036)	(0.051)
IV, White	-0.030	-0.030	0.000	0.050
	(0.201)	(0.120)	(0.117)	(0.163)
P-value, test of diff.	0.470	0.520	0.965	0.273
CCM, Minority	3.59	1.98	2.35	2.20
CCM, White	3.52	2.00	2.31	2.03
N	43019	43019	36300	36300

Notes: Spells defined as consecutive quarters with earnings from same employer. Time to first quarter conditions on spells > 0. CM shows control means; CCM shows control complier means. Regression includes baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.21: Employment and Earnings Effects by Race/Ethnicity, ITT

	Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, White	0.0048	-0.0017	-0.003	-69.27	-160.38	-227.40
	(0.0114)	(0.0122)	(0.0096)	(144.83)	(218.91)	(328.78)
ITT, Black	0.0078	0.0047	0.0052	26.81	90.36	128.61
	(0.0064)	(0.0063)	(0.0050)	(68.65)	(108.67)	(158.35)
ITT, Hispanic	0.0162**	0.0066	0.0118*	189.00**	209.86	387.14*
	(0.0077)	(0.0077)	(0.0064)	(88.21)	(139.93)	(206.26)
ITT, Asian	0.0250**	0.0024	0.0144	-12.56	82.98	59.91
	(0.0121)	(0.0121)	(0.0101)	(116.81)	(206.84)	(286.70)
ITT, Other	0.0103	0.0330*	0.0077	106.66	446.60	539.64
	(0.0194)	(0.0192)	(0.0161)	(206.74)	(322.74)	(467.47)
P-value, all equal	0.681	0.646	0.676	0.462	0.528	0.490
CM, White	0.7518	0.6949	0.8510	3754	5702	9457
CM, Black	0.7146	0.7434	0.8568	3575	5937	9512
CM, Hispanic	0.6868	0.7182	0.8297	3726	6365	10092
CM, Asian	0.6430	0.6751	0.8069	3010	5256	8266
CM, Other	0.6849	0.705	0.8295	3538	5591	9128

Notes: N = 5,366 White, N = 17,636 Black, N = 12,427 Hispanic, N = 5,578 Asian, and N = 2,012 Other youth with 390 observations dropped due to missing race/ethnicity. Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means. P-value from test of null hypothesis that all treatment effects are equal across groups. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.22: Employment and Earnings Effects by Race/Ethnicity, IV

	First Stage	Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
IV, White	0.2972***	0.0156	-0.0071	-0.011	-241.10	-562.53	-796.71
	(0.0088)	(0.0385)	(0.0412)	(0.0323)	(488.46)	(738.11)	(1108.55)
IV, Black	0.4039***	0.0194	0.0117	0.0128	74.25	237.82	338.74
	(0.0052)	(0.0157)	(0.0155)	(0.0124)	(169.75)	(268.69)	(391.44)
IV, Hispanic	0.4152***	0.0390**	0.0158	0.0285*	453.07**	500.75	926.35*
	(0.0062)	(0.0186)	(0.0185)	(0.0154)	(212.34)	(336.73)	(496.50)
IV, Asian	0.4830***	0.0518**	0.0051	0.0298	-12.56	195.34	159.76
	(0.0094)	(0.0250)	(0.0250)	(0.0210)	(241.70)	(428.34)	(593.51)
IV, Other	0.3926***	0.026	0.0838*	0.0192	283.16	1152.15	1398.40
	(0.0155)	(0.0496)	(0.0492)	(0.0411)	(528.36)	(826.74)	(1195.76)
P-value, all equal	0.000	0.811	0.658	0.780	0.494	0.574	0.534
CCM, White		0.753	0.714	0.865	4406	6610	11008
CCM, Black		0.723	0.764	0.867	3729	6294	9996
CCM, Hispanic		0.681	0.722	0.826	3760	6397	10184
CCM, Asian		0.630	0.667	0.790	3128	5168	8318
CCM, Other		0.691	0.694	0.846	3919	5291	9247

Notes: N = 5,366 White, N = 17,636 Black, N = 12,427 Hispanic, N = 5,578 Asian, and N = 2,012 Other youth with 390 observations dropped due to missing race/ethnicity. Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CDM shows control complier means. P-value from test of null hypothesis that all treatment effects are equal across groups. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.23: Employment and Earnings Effects for Male and Female Youth

	First Stage	Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Male		0.0045	0.0097	0.004	44.40	154.23	197.37
		(0.0066)	(0.0067)	(0.0056)	(66.40)	(108.86)	(158.56)
ITT, Female		0.0188***	0.0029	0.0109***	71.69	77.20	147.27
		(0.0052)	(0.0052)	(0.0041)	(62.94)	(98.67)	(144.19)
P-value, test of diff.		0.09	0.423	0.326	0.765	0.600	0.815
CM, Male		0.658	0.659	0.802	3015	4997	8012
CM, Female		0.733	0.766	0.870	4000	6684	10684
IV, Male	0.3962***	0.0113	0.0245	0.01	118.63	398.51	520.97
	(0.0051)	(0.0166)	(0.0168)	(0.0142)	(167.44)	(274.52)	(399.62)
IV, Female	0.4106***	0.0457***	0.0071	0.0263***	179.64	197.12	369.99
	(0.0044)	(0.0128)	(0.0127)	(0.0100)	(153.23)	(240.24)	(351.06)
P-value, test of diff.	0.031	0.101	0.410	0.347	0.788	0.581	0.776
CCM, Male		0.675	0.667	0.812	3243	5195	8434
CCM, Female		0.713	0.773	0.862	4079	6858	10944

Notes: Notes: N = 18,539 male youth and N = 24,870 female youth. Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.24: Employment and Earnings Effects by Self-Reported High School Enrollment

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, High School		0.0149***	0.0107**	0.0095**	66.8671*	99.95	170.2714*
		(0.0050)	(0.0049)	(0.0041)	(39.50)	(66.42)	(94.49)
ITT, Non-High School		0.0059	-0.0091	0.003	44.36	137.33	180.36
		(0.0070)	(0.0075)	(0.0056)	(141.85)	(217.48)	(324.66)
P-value, test of diff.		0.291	0.026	0.348	0.879	0.869	0.976
CM, High School		0.662	0.696	0.822	2492	4473	6965
CM, Non-High School		0.822	0.795	0.898	6934	10566	17499
	First Stage						
IV, High School	0.3992***	0.0375***	0.0267**	0.0237**	174.5338*	261.27	435.8011*
	(0.0038)	(0.0124)	(0.0122)	(0.0102)	(98.93)	(166.25)	(236.77)
IV, Non-High School	0.4204***	0.0137	-0.0214	0.0071	94.97	356.81	451.78
	(0.0067)	(0.0166)	(0.0177)	(0.0133)	(336.41)	(515.82)	(769.45)
P-value, test of diff.	0.006	0.253	0.026	0.323	0.821	0.86	0.984
CCM, High School		0.658	0.698	0.823	2565	4549	7114
CCM, Non-High School		0.81	0.816	0.891	7058	10757	17815

Notes: N = 32,703 high school youth and N = 10,706 non-high school youth (with 1 observation missing education coded as non-high school). Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.25: Employment and Earnings Effects by Age

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Under 18		0.0136**	0.0086	0.0085*	45.02	59.21	104.14
		(0.0055)	(0.0054)	(0.0045)	(37.98)	(66.43)	(92.54)
ITT, 18 and Over		0.0109*	0.0003	0.0066	87.43	191.01	298.05
		(0.0060)	(0.0063)	(0.0048)	(107.02)	(165.19)	(245.86)
P-value, test of diff.		0.745	0.317	0.773	0.709	0.459	0.46
CM, Under 18		0.645	0.686	0.815	2145	3985	6129
CM, 18 and Over		0.798	0.780	0.886	6067	9395	15462
	First Stage						
IV, Under 18	0.4027***	0.0337**	0.0213	0.0212*	115.81	155.71	269.76
	(0.0042)	(0.0136)	(0.0134)	(0.0112)	(94.22)	(164.67)	(229.81)
IV, 18 and Over	0.4072***	0.0266*	0.0006	0.0157	218.86	485.31	721.01
	(0.0055)	(0.0148)	(0.0155)	(0.0119)	(262.60)	(405.77)	(603.82)
P-value, test of diff.	0.510	0.724	0.313	0.738	0.712	0.451	0.485
CCM, Under 18		0.647	0.695	0.821	2272	4198	6472
CCM, 18 and Over		0.783	0.786	0.876	6211	9516	15710

Notes: N = 42,409. Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.26: Employment and Earnings Effects by Neighborhood: Above/Below Median in Opportunity Insights Upward Mobility Ranking

	First Stage	Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Above Median		0.0111*	0.0016	0.0032	64.18	55.26	115.83
		(0.0058)	(0.0059)	(0.0048)	(66.53)	(106.74)	(155.19)
ITT, Below Median		0.0143**	0.0100*	0.0126***	56.21	164.90	220.49
		(0.0058)	(0.0057)	(0.0047)	(63.27)	(100.37)	(147.04)
P-value, test of diff.		0.694	0.307	0.159	0.931	0.454	0.624
CM, Above Median		0.706	0.711	0.840	3537	5898	9435
CM, Below Median		0.696	0.729	0.841	3621	6029	9650
IV, Above Median	0.3903***	0.0285*	0.004	0.0082	172.51	152.86	321.62
	(0.0047)	(0.0150)	(0.0152)	(0.0123)	(170.54)	(273.50)	(397.67)
IV, Below Median	0.4183***	0.0343**	0.0239*	0.0299***	138.76	401.58*	539.28
	(0.0047)	(0.0138)	(0.0137)	(0.0112)	(151.10)	(239.64)	(351.08)
P-value, test of diff.	0.000	0.775	0.331	0.193	0.882	0.494	0.682
CCM, Above Median		0.693	0.714	0.837	3645	6015	9664
CCM, Below Median		0.700	0.741	0.844	3805	6293	10099

Notes: N = 42,408 (1 observation is missing zip code). Above/below median defined as the within-sample median of the Opportunity Insights "upward mobility" index: the average percentile rank for children whose parents were in the 25th percentile of the national income distribution. We map Census tract-level data onto study participant zip code, see text for details. Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.27: Employment and Earnings Effects by Pre-SYEP Work Experience Status

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Ever Worked		0.0177***	0.0065	0.0117***	72.08	184.32	253.99
		(0.0050)	(0.0051)	(0.0038)	(77.80)	(118.37)	(176.63)
ITT, Never Worked		0.0065	0.005	0.0035	39.60	10.91	58.45
		(0.0068)	(0.0067)	(0.0058)	(38.28)	(75.09)	(100.48)
P-value, test of diff.		0.183	0.862	0.237	0.708	0.216	0.336
CM, Ever Worked		0.793	0.790	0.895	5074	8000	13073
CM, Never Worked		0.589	0.635	0.774	1749	3471	5220
	First Stage						
IV, Ever Worked	0.4120***	0.0431***	0.0157	0.0280***	175.05	448.45	622.09
	(0.0045)	(0.0121)	(0.0123)	(0.0092)	(188.64)	(287.28)	(428.80)
IV, Never Worked	0.3951***	0.0164	0.0127	0.0086	106.76	46.19	145.86
	(0.0049)	(0.0173)	(0.0170)	(0.0147)	(96.63)	(189.66)	(254.18)
P-value, test of diff.	0.011	0.207	0.885	0.263	0.747	0.242	0.34
CCM, Ever Worked		0.775	0.794	0.883	5209	8057	13267
CCM, Never Worked		0.601	0.647	0.789	1920	3846	5772

Notes: N = 23,731 youth with work experience prior to the SYEP summer and N = 19,678 youth who never worked prior to the SYEP summer. Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.28: Letter Information and Application Behavior for Treatment Group by Subgroup

	Has Letter	Average Rating	Applied to Our Job	Submitted Letter
White	0.296	6.09	0.073	0.222
Minority	0.420	5.66	0.083	0.158
Black	0.404	5.54	0.087	0.167
Hispanic	0.416	5.68	0.071	0.130
Asian	0.483	5.85	0.121	0.200
Male	0.396	5.62	0.077	0.162
Female	0.410	5.75	0.086	0.167
In High School	0.399	5.64	0.084	0.198
Not in HS	0.421	5.85	0.077	0.053
Under 18	0.403	5.64	0.084	0.226
18 and Over	0.407	5.79	0.079	0.052
Above Median in OI Rank	0.390	5.80	0.077	0.167
Below Median in OI Rank	0.418	5.60	0.086	0.163
Never Employed Pre-SYEP	0.395	5.61	0.077	0.214
Ever Employed Pre-SYEP	0.412	5.77	0.086	0.128
High Rating	0.753	6.04	0.093	0.250
Low Rating	0.307	3.92	0.074	0.167

Notes: Means shown for treatment group only, N = 21,714 (except for high/low rating, which is limited to those with a rating, N = 14,400). Average rating conditional on being sent a letter, N = 8,780; application probability conditional on being invited to apply, N = 2,000 (1,346 for rating categories); and submission probability conditional on applying, N = 164 (116 for rating categories). Median OI Rank is the within-sample median of the Opportunity Insights "upward mobility" percentile rank. All differences in having a letter and in average ratings between two groups (i.e., White/Minority, Male/Female, High School/Not in HS, Under/Over 18, Above/Below median OI rank, Never/Ever Employed Pre-SYEP, and High/Low Ratings) are statistically different except for having a letter between those under and over 18. None of the differences in application or letter submission rates are significantly different except for the high school and age differences in submitting the letter.

Table A.29: Baseline Characteristics, Unopened Surveys versus Main Control Group

	Unopened Surveys	Control	Test of Difference
N	25813	21695	
Age	17.24	17.17	0.002
Male	0.427	0.427	0.894
Black	0.437	0.409	0.000
Hispanic	0.246	0.289	0.000
Asian	0.082	0.129	0.000
White	0.188	0.124	0.000
Other Race	0.047	0.049	0.746
In High School	0.746	0.755	0.014
HS Graduate	0.050	0.044	0.003
In College	0.174	0.173	0.677
Not in UI Data	0.011	0.009	0.039
Never Employed Pre-SYEP	0.429	0.450	0.000
Ever Worked, Year -4	0.170	0.153	0.000
Earnings, Year -4	333	318	0.443
Ever Worked, Year -3	0.293	0.267	0.000
Earnings, Year -3	616	585	0.174
Ever Worked, Year -2	0.459	0.437	0.000
Earnings, Year -2	1113	1072	0.182
Ever Worked, Year -1	0.962	0.966	0.042
Earnings, Year -1	2379	2379	0.775
No Education Match	0.185	0.126	0.000
In HS Sample	0.409	0.454	0.000
Joint F-test	$F(24, 45597) = 35.48$, $p=0$		

Notes: Table tests difference of means between all youth in unopened surveys (excluded from our main sample) and our control group (on an opened survey). Test of difference controls for cohort indicator and uses cluster-robust standard errors. 496 youth are missing race/ethnicity.

Table A.30: Outcomes, Unopened Surveys versus Main Control Group

	Unopened Surveys	Control	Test of Difference
Panel A: Labor Market Outcomes			
N	25813	21695	
Employment Y1	0.715	0.701	0.001
Employment Y2	0.715	0.720	0.256
Employment Cumulative	0.843	0.841	0.417
Earnings Y1	3665	3579	0.155
Earnings Y2	6190	5964	0.009
Earnings Cumulative	9855	9543	0.020
Joint F-test, Employment Outcomes	F(5, 45597) = 5.96, p=0		
Panel B: Education Outcomes			
N	10564	9857	
Enrolled Y1	0.934	0.946	0.000
Days Enrolled Y1	146.8	150.0	0.000
Share Days Attended Y1	0.808	0.829	0.000
GPA Y1	79.03	80.13	0.000
On-time Graduation	0.759	0.791	0.000
Ever Graduated	0.804	0.836	0.000
Grad or Still in School	0.862	0.887	0.000
On-time College	0.665	0.697	0.000
Joint F-test, Y1 and Ever Graduated Outcomes	F(5, 18038) = 91.95, p=0		

Notes: Table tests difference of means between all youth in unopened surveys (excluded from our main sample) and our control group (on an opened survey), separately for employment outcomes and subset of youth in expected HS sample. N = 19,239 for graduation tests and N = 14,543 for college test. To avoid using the smallest available sample and highly correlated outcomes for joint F-test, the education joint test includes 4 high school outcomes and an ever graduated indicator. Test of difference controls for cohort indicator and uses cluster-robust standard errors.

Table A.31: Labor Market Effects, On Any Survey

Year	1	2	Cumulative
Panel A: Employment			
ITT	0.0062*	0.0034	0.0029
	(0.0032)	(0.0033)	(0.0027)
CM	0.707	0.719	0.843
Sent Letter (IV)	0.0244*	0.0136	0.0111
	(0.0128)	(0.0129)	(0.0104)
CCM	0.704	0.729	0.849
Panel B: Earnings, Winsorized at 99.5th Percentile			
ITT	23.34	65.37	89.04
	(36.70)	(58.97)	(85.73)
CM	3619	6052	9671
Sent Letter (IV)	92.75	257.2	350.06
	(144.43)	(232.05)	(337.29)
CCM	3791	6186	9977
Panel C: Log(Earnings + 1)			
ITT	0.043*	0.034	0.028
	(0.026)	(0.028)	(0.023)
CM	5.66	6.08	7.35
Sent Letter (IV)	0.168*	0.135	0.109
	(0.102)	(0.110)	(0.092)
CCM	5.71	6.20	7.47

Notes: N = 69,222. Sample includes all youth on any survey, regardless of whether any supervisor opened the survey. Earnings winsorization recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.32: Education Effects, On Any Survey & In Expected HS Sample

	Ever Enrolled Y1	Days Enrolled Y1	Share Days Present Y1	GPA Y1	On-time Graduation	Ever Graduated	Graduated or Still Attending	On-time College
ITT	-0.0008	-0.0625	0.001	-0.0576	-0.0037	0.0014	-0.0005	0.0006
	(0.0025)	(0.4527)	(0.0023)	(0.0804)	(0.0033)	(0.0033)	(0.0031)	(0.0050)
CM	0.942	148.9	0.822	79.76	0.781	0.825	0.879	0.687
N	30278	30278	30278	27868	28491	28491	28491	21533
Sent Letter (IV)	-0.0025	-0.2313	0.0037	-0.2145	-0.0134	0.0053	-0.0015	0.0028
	(0.0092)	(1.6760)	(0.0084)	(0.2900)	(0.0122)	(0.0120)	(0.0112)	(0.0177)
CCM	0.955	152.6	0.848	81.66	0.831	0.859	0.913	0.719
N	30278	30278	30278	27868	28491	28491	28491	21533

Notes: Sample includes all youth on any survey, regardless of whether any supervisor opened the survey, if they were expected to be observed based on pre-program grade of enrollment (see text for details). On-time graduation equals 1 for public school, non-charter students who graduate within 4 years and do not transfer to GED programs or other districts. Ever graduated includes any 5th- or 6th-year graduation for the same group if it occurs during our data. Both measures include 8th–12th graders. College enrollment is only measured within 6 months after a student's on-time graduation date, regardless of graduation status. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.33: Labor Market Effects, Alternative Covariates

Year	1	2	Cumulative	1	2	Cumulative
	No Covariates			All Covariates		
	Panel A: Employment					
ITT	0.0115***	0.0048	0.0070**	0.0124***	0.0058	0.0077**
	(0.0044)	(0.0043)	(0.0035)	(0.0041)	(0.0041)	(0.0034)
CM	0.701	0.72	0.841	0.701	0.72	0.841
Sent Letter (IV)	0.0284***	0.0119	0.0172**	0.0306***	0.0143	0.0190**
	(0.0108)	(0.0106)	(0.0086)	(0.0101)	(0.0102)	(0.0083)
CCM	0.7	0.731	0.843	0.698	0.728	0.841
	Panel B: Earnings, Winsorized at 99.5th Percentile					
ITT	41.81	102.64	144.46	55.14	109.8	164.93
	(54.62)	(82.90)	(126.80)	(45.77)	(73.12)	(106.75)
CM	3579	5964	9543	3579	5964	9543
Sent Letter (IV)	103.41	253.84	357.25	136.33	271.48	407.81
	(135.05)	(204.98)	(313.49)	(113.11)	(180.73)	(263.85)
CCM	3780	6190	9970	3747	6172	9919
	Panel C: Log(Earnings + 1)					
ITT	0.084**	0.049	0.066**	0.093***	0.058*	0.073**
	(0.036)	(0.038)	(0.032)	(0.0330)	(0.0350)	(0.0300)
CM	5.61	6.08	7.33	5.61	6.09	7.33
Sent Letter (IV)	0.208**	0.122	0.162**	0.229***	0.144*	0.181**
	(0.089)	(0.093)	(0.079)	(0.0810)	(0.0870)	(0.0730)
CCM	5.67	6.21	7.41	5.64	6.19	7.40

Notes: N = 43,409. Left panel shows results with no coviarates other than cohort indicator. Right panel uses all available covariates (see text) rather than post-double-selection LASSO-selected covariates that are used in the main results. Winsorization in Panel B recodes each quarter's highest earnings to the 99.5th percentile of all quarterly earnings before summing across years. CM shows control means; CCM shows control complier means. Regressions include cohort indicator only. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.34: Education Effects, Alternative Covariates

	Ever Enrolled Y1	Days Enrolled Y1	Share Days Present Y1	GPA Y1	On-time Graduation	Ever Graduated	Graduated or Still Attending	On-time College
	Panel A: No Covariates							
ITT	-0.0033	-0.8474	-0.0038	-0.4625**	-0.0137**	-0.0074	-0.0075	-0.0118
	(0.0033)	(0.6344)	(0.0038)	(0.1842)	(0.0060)	(0.0055)	(0.0047)	(0.0078)
CM	0.946	150.0	0.829	80.13	0.791	0.836	0.887	0.697
N	19714	19714	19714	18237	18537	18537	18537	13999
Sent Letter (IV)	-0.0081	-2.0549	-0.0093	-1.1128**	-0.0329**	-0.0178	-0.018	-0.0272
	(0.0080)	(1.5393)	(0.0092)	(0.4445)	(0.0144)	(0.0131)	(0.0112)	(0.0180)
CCM	0.961	154.4	0.861	82.56	0.851	0.882	0.930	0.749
N	19714	19714	19714	18237	18537	18537	18537	13999
	Panel B: All Covariates							
ITT	-0.0018	-0.3201	0.0015	-0.1354	-0.0066	-0.0022	-0.0032	0.0004
	(0.0030)	(0.5425)	(0.0028)	(0.0983)	(0.0041)	(0.0040)	(0.0037)	(0.0062)
CM	0.9455	150.0	0.8285	80.13	0.7914	0.8358	0.8871	0.6968
N	19714	19714	19714	18237	18537	18537	18537	13999
Sent Letter (IV)	-0.0043	-0.7724	0.0036	-0.3243	-0.0156	-0.0052	-0.0076	0.001
	(0.0072)	(1.3079)	(0.0067)	(0.2353)	(0.0097)	(0.0094)	(0.0089)	(0.0141)
CCM	0.957	153.1	0.848	81.77	0.834	0.869	0.92	0.721
N	19714	19714	19714	18237	18537	18537	18537	13999

Notes: Analysis is conducted on all those expected to be observed based on pre-program grade of enrollment (see text for details). Panel A shows results with no coviarates other than cohort indicator. Panel B uses all available covariates (see text) rather than post-double-selection LASSO-selected covariates that are used in the main results. On-time graduation equals 1 for public school, non-charter students who graduate within 4 years and do not transfer to GED programs or other districts. Ever graduated includes any 5th- or 6th-year graduation for the same group if it occurs during our data. Both measures include 8th–12th graders. College enrollment is only measured within 6 months after a student's on-time graduation date, regardless of graduation status. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$