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)

September 17, 2021

NOTE TO NBER ORGANIZERS: Please do not cite, circulate, or post this draft, as it is still a bit rough and we do not yet have permission from data providers to make it public (though we can present results). We note that because our UI data access is limited to when we are on Penn's campus, and Penn is still restricting campus access due to Covid-19, we have not been able to complete all the analyses we have planned. We plan to push a few more things forward prior to the fall meeting.

Abstract

Youth employment is near historic lows, and racial gaps persist. We test whether informational frictions contribute to these problems by running a large-scale field experiment—including over 43,000 youth in New York City—aimed at providing credible information on young applicants. 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, then send these letters to nearly 9,000 youth. Being sent a letter generates a 3 percentage point (4.4 percent) increase in employment the following year, with both employment and earnings increases persisting over the 2-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 a supply-side response (increased job search, motivation, or confidence); effects appear due to demand-side responses. Labor market benefits accrue primarily to racial and ethnic minorities, suggesting frictions may contribute to racial employment gaps. But improved employment also decreases on-time high school graduation, especially among minorities. Expanding the availability of credible signals about young workers, especially non-White applicants who are no longer in 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. We are particularly grateful to Julia Breitman at DYCD for all of her help along the way. 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 continually lags farther behind adult employment than can be explained by school engagement, and it consistently recovers much more slowly in the wake of major shocks. Even a decade after the Great Recession, youth employment rates in the summer—when teenagers are most likely to be working—were still hovering near their sixty-year low (DeSilver 2021). And the Covid-19 pandemic appears to have disproportionately harmed youth labor market prospects, in part because of the concentration of young people employed in the food service and retail sectors (Inanc 2020; Kochhar and Barroso 2020). Employment gaps are particularly bleak for minority youth; unemployment and disconnection rates are 70–80 percent higher for Black and Latino youth than for white youth (Bureau of Labor Statistics 2020; Spievack and Sick 2019), a pattern of disproportionate impacts that is not new (Sum et al. 2014). That young people, and minority youth in particular, have such difficulty securing employment is quite troubling given the range of evidence suggesting that labor force attachment during adolescence and young adulthood affects employment and wages for decades (Baum and Ruhm 2016; Kahn 2010; Neumark 2002; Oreopoulos, Wachter, and Heisz 2012).

Half a century of active labor market programs have focused on relatively costly efforts to improve human capital or provide job search assistance for young people struggling in the labor market. Though sometimes successful (see discussion in Katz et al. 2020), the frequency with which youth training programs fail to improve employment raises the possibility that young people may face other important frictions in the labor market. Their short or nonexistent work histories may leave little way to credibly signal future productivity. Employers may view participation in training programs as a negative signal itself, or employers may engage in statistical or animus-driven discrimination based on age or race. Alternatively, on the supply side, young people may lack the access to jobs, knowledge, or motivation to convert their experience into better employment outcomes. Although there is some evidence that informational frictions matter in an online marketplace and for women in South Africa (Pallais 2014; Abel, Burger, and Piraino 2020), we know remarkably little about the role such frictions play in U.S. labor markets for young people, nor whether such frictions disproportionately affect racial and ethnic minorities in a way that could contribute to persistent racial gaps.

In this paper, we explore the potential role of informational frictions in the youth labor market by testing an intervention designed to mitigate them. We run a large-scale field experiment with over 43,000 youth in New York City (NYC), partnering with the NYC Summer Youth Employment Program (SYEP), which employs city youth to work over the summer

months. We provide a random subset of youth participants with letters of recommendation from their supervisors in the program. 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 youth participants into full-text letters of recommendation. When supervisors agree to produce a letter of recommendation and provide high-enough ratings of youth, treatment youth receive a digital copy and five hard copies of the letter.

We follow treatment and control youth in administrative data for two years after treatment to explore 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. We measure employment and earnings using data from the New York State Department of Labor. We observe educational outcomes 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 their summer 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 either future employment or educational attainment (Davis and Heller 2020; Gelber, Isen, and Kessler 2016; Modestino 2019).[1]

Our results suggest a sizable impact of the letter of recommendation intervention. We find that being sent a letter of recommendation increases the likelihood that a youth is employed by over 3 percentage points in the year after receiving the letter, a 4.4 percent increase relative to the 70 percent of their control counterparts who work.[2] Employment effects appear to persist over time, with impacts remaining positive over the cumulative 2-year follow-up period; youth who are sent a letter are 1.9 percentage points more likely to have a job over the next two years, a roughly 2.3 percent increase on a base of 84 percentage points. Similarly, cumulative earnings are at least 4 percent higher for those sent the letter (p=0.12), with different adjustments for skewness suggesting considerably bigger and more

1. This is in spite of the fact that one of the program goals of SYEPs is to improve the long-term employment outcomes of participants (Gelber, Isen, and Kessler 2016).

2. This effect is 250% as large as estimates of the effect of the summer program itself on employment in the year after the program. The program has been shown to increase employment by 1.2 percentage points in that year by encouraging youth to participate in SYEP again (Gelber, Isen, and Kessler 2016).

statistically significant effects ($p<0.05$).

That simply providing a few pieces of paper improves employment and earnings suggests an important role for information frictions in the youth labor market. Such frictions could arise on either the supply or the demand side. On the supply side, if youth lack confidence or do not realized that their summer employment may make them more marketable to future hiring managers, then they may decide not to apply to jobs that would be a good fit for their skills or not to market their summer experience. On the demand side, if employers do not understand what skills youth develop in their summer jobs or if they engage in statistical discrimination against youth—or minority youth in particular—they might discount the human capital of young applicants unless provided with a signal of their ability. Our treatment could potentially mitigate both sets of frictions. 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 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) to apply for a short-term online job working for us. The applications we received confirmed that youth use the letter: Treatment youth were 250% 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). Since only about 40 percent of letter recipients actually use them on our application, our main IV effect of being sent a letter likely understates the effect of the letters once employers see them. Back-of-the-envelope calculations—which involve scaling our employment estimates using data on the rate that letters are submitted as part of our job application—suggest that actually using the letter is associated with a 15 percent increase in employment in the first year and almost $1,400 in additional earnings (also about 15 percent) over two years.

Importantly, 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 neither increased motivation to apply to jobs nor increased confidence are driving our employment effects.[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. This finding

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

highlights the possible presence of demand-side frictions in the youth labor market, and it implies that finding additional ways to communicate credible information about youth applicants to potential employers may help youth succeed in the labor market.

Indeed, heterogeneous treatment effect analysis of our employment and earnings data reveal that youth who perform better in their summer jobs see a larger increase in employment and earnings from a letter of recommendation, suggesting an important role for information transmission in the letters. We also find that the labor market benefits accrue primarily to minority youth, suggesting that this group may face larger frictions in the youth labor market market.

While our effects are large, they are consistent with past research that suggests that providing even a small amount of information about job-seekers can be quite powerful. Moderately positive feedback in an online marketplace significantly improves an applicant's later employment outcomes in the marketplace (Pallais 2014). On fictitious applications from audit studies, employers show less discrimination when they have more information about people (Agan and Starr 2018; Kaas and Manger 2012). And in the work most similar to ours, when job seekers in South Africa were encouraged to seek recommendation letters, the letters helped close gender gaps by providing better information on high-quality applicants, including improving self-reported measures of employment for women after 3 months (Abel, Burger, and Piraino 2020). Both here and in our context, one might worry that the large effects could be driven by employers *incorrectly* updating beliefs (e.g., based on their priors about who submits recommendation letters). Our longer follow-up with administrative data can assess this possibility, since encouraging bad matches should lead to more job transitions. In fact, we find no evidence of increased churn as measured by the number of job spells or the length of each spell. If anything, spells seem to get longer for treatment youth, suggesting the letters are helping workers and employers make more successful job matches.

We collected education information in addition to employment data for two reasons. First, letters could have a direct effect if shown to teachers or guidance counselors: other evidence has shown that teachers' and 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 more than 20 hours is harmful but working a small amount has little effect (Buscha et al. 2012; Staff, Schulenberg, and Bachman 2010; Monahan, Lee, and Steinberg 2011; Baum and Ruhm 2016; Ruhm 1997). Yet 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.

In fact, we find that the letters of recommendation are not just associated with higher employment and earnings, but also seem to slow down high school graduation. Being sent a letter decreases on-time, 4-year graduation by 3.6 percentage points (4 percent). It is less clear whether they decrease final graduation outcomes on average, since point estimates for measures that include 5- and 6-year graduation are less than half as big and not statistically significant on average. However, consistent with our employment results, this decrease in on-time high school graduation is concentrated among minority youth, whose decline in overall school persistence (measuring whether someone has either graduated or is still attending school after 2 years) is marginally significant. Together, these results suggest that increasing labor market participation on the margin does substitute for school for young workers. They also suggest that the welfare implications of efforts to reduce informational frictions in the labor market targeted at young people may depend on whether or not the young people are still in school.

Our study shows that information frictions do prevent young people, especially minority youth, from working. 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—especially those who have finished their education. One concern about more widespread use of credible information signals could be displacement; youth with these signals might just displace those without them (though 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). The potential general equilibrium effects are worth considering as part of efforts to scale up such programs, though policymakers may value the distributional changes or efficiency gains from better matches even if there were no net change in employment. It is also 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. Regardless, the role of informational frictions in depressing employment and earnings for young people, as well as potentially limiting the impact of programs designed to improve their skills, is now clear.

2 Setting, Experiment, and Data

2.1 Setting

We partner with the New York City Summer Youth Employment Program (NYC SYEP) and implement our experiment with youth who participated in the summer of 2016 or the

summer of 2017.[4] In these program years, the NYC SYEP provided youth with 6 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 (NYS) minimum wage ($11.00 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 NYS (see *SYEP Annual Summary* 2017).

Youth apply to the NYC SYEP through providers throughout the five boroughs. These providers accept applications from youth, and the SYEP lottery—which determines whether a youth is offered a spot in the program—is run at the provider level.[5] Youth who are offered a spot in the program (e.g., because they win the lottery) can choose whether or not to participate. Providers maintain relationships with employers and aim to match participants with job opportunities based on their age, experience, and interests.

2.2 Letter of Recommendation Experiment

We received SYEP data from from the NYC Department of Youth and Community Development (DYCD)—which administers the NYC SYEP—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 likely supervisor or supervisors for the youth at that work site.[6] Using these data, we limited our sample in several

4. The NYC SYEP is administered by the NYC Department of Community Development (DYCD). Since a post-Great Recession minimum enrollment of 29,416 youth, enrollment grew steadily to nearly 70,000 youth in 2017. It is oversubscribed in each year; in 2017, there were over 147,000 applications submitted for the 70,000 available jobs. For most of the program slots, participation is determined by lottery. In addition, there are two groups of youth who are not selected by lottery. The first are "Vulnerable Youth" are defined by DYCD as youth who are: (a) involved in the criminal justice system, (b) homeless or have run away from home, (c) currently in or aging out of foster care, (d) part of families receiving preventive services through NYC's Administration for Children's Services. The second are youth participating in "Ladders for Leaders," a program that selects that selects high-achieving youth by application and aims to provide them with jobs that match their skills and interests.

5. While the lottery is centralized, the spots are allocated within a provider, so applicants at different providers may have slightly different chances of winning the lottery.

6. 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

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.[7] In total, this left a sample of 69,224 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 administer the survey, 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). The email inviting each supervisor to participate explained a bit about 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.I). 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 the survey. 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.II). Once the supervisor confirmed being a supervisor, they were shown the list of treatment youth linked to them in the DYCD data, listed alphabetically by last name.[8] 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.II for screenshots of the survey). If they were willing, they were also asked if

names of others who might have supervised youth so we could include them in the letter of recommendation program as well.

7. Note that we randomly assigned treatment and control status prior to making these sample limitations, so that we could ensure that neither the treatment or control group exceeded the 30-person-per-survey limit. Since youth were randomly selected to be excluded, random assignment is still only a function of random variables.

8. 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.

they wanted to include their contact information on the letter of recommendation to serve as a future reference. 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.III). They were told that these youth would not be included in the letter of recommendation program.

A total of 5,760 supervisors (1,333 in 2016 and 4,427 in 2017, about 40 percent of all supervisors we emailed) opened the survey and made it to the screen where they could start rating treatment youth. In total, 29,887 youth, or 69 percent of youth on a started survey, were given an overall rating by employers (23,909 in 2017 and 5,978 in 2016).

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. That is, 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 attributes to a potential paragraph. If the supervisor rated the youth positively enough on enough attributes to construct a paragraph, the paragraph was included in the letter. As long as 2 paragraphs could be included, the letter was generated for the youth. This procedure ensured that any letters of recommendation generated by our survey tool were positive (i.e., had enough positive things to say about the youth to justify providing a letter). 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, 8,780 letters were generated and sent to youth (1,805 in 2016; 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; 6,720 in 2017).[9] In addition, five physical copies of the letters were sent via USPS to each youth along with a cover letter providing context and suggested uses for the letter (see Appendix Figure A.IV for a sample cover letter; similar text was sent to youth via email along with the link to the soft copy of the letter).[10] All

9. 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.

10. Of the 8,780 letters mailed to youth, 138 were returned as undeliverable.

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 mechanism through which letters of recommendation might impact labor market outcomes of treatment youth, for a subset of the youth in our data we: advertised a job, solicited job applications, and hired youth ourselves. We started with a job listing for a one-time, remote paid work assignment. We 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.V).

Youth who clicked the link in the email were taken to a job application that asked the youth a few standard contact, background, and employment experience questions. It 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.VI).[11] 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 to check a box on the application to be considered for a more selective, higher-paying position ($18/hour). The application told them explicitly that being considered for the more selective

11. We intentionally avoided explicitly mentioning a letter of recommendation to see if youth in our study would choose to upload a letter without an explicit mention of one. 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.

position would not affect their chances at being selected for the regular job.

All applicants who submitted an application that included their name and email address were hired.[12] The job itself was an online survey of multiple choice questions. These questions asked about youth experiences job-seeking and preparing for 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.[13] Workers were instructed to finish everything they could within a two hour time frame. All youth who initiated the job-task (n=296) were paid for two hours of work via a mailed gift card.

2.4 SYEP Administrative Data

Administrative data from the NYC SYEP comes from the NYC Department of Youth and Community Development (DYCD), which administers 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, family size, and citizenship 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 these self-reported categories, making the classifications mutually exclusive (i.e., "White" only captures non-Hispanic Whites). We also received information on the providers through which the youth applied to SYEP (see discussion in Section 2), 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 construct the letter of recommendation survey software.

2.5 NYS Department of Labor Data

Earnings and employment data were obtained 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 data includes 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

12. 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 standard job.

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

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 ($\beta = 0.002, p = 0.13$).

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 data for both study cohorts after letters were sent to SYEP participants in our treatment group.[14]

2.6 NYC Department of Education Data

Education data come from the NYC Department of Education (DOE). 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. 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 ($\beta = -0.003, p = 0.359$). At the request of the data provider, we keep DOE data in a separate file that excludes citizenship status, so that education outcomes are never linked to the self-reported citizenship status from the SYEP application.[15]

3 Method of Analysis

This section discusses how we perform the analysis in this paper. In Section 3.1, we describe our sample definitions, showing treatment-control balance where appropriate. In Section 3.2, we describe our outcomes of interest. In Section 3.3, 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.[16]

3.1 Sample Definitions

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

14. 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.

15. 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.

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

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 that they supervised youth that summer).[17]

While we needed to randomize all SYEP participants to treatment and control (because we needed a treatment status to build the personalized surveys for supervisors and to ensure that a youth who appeared on multiple surveys had the same treatment status on each one), we pre-specified this smaller group as a key sample of interest, since the non-response of their supervisors means youth outside of our main sample did not have any chance of receiving a letter of recommendation, regardless of their treatment status. This kind of non-compliance mechanically reduces statistical power by introducing noise.

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 began the survey they had no way of knowing which youth would be included in the survey and which youth would be in the treatment or control groups). As expected, Table 1 shows that our main sample is balanced across treatment and control youth. It is worth noting, however, that the treatment effect of receiving a letter of recommendation in our main sample is estimated on the population of youth whose supervisors: 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 treatment effect estimated on 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. Tables A.XVI through A.XVIII show that youth who were on surveys who were not opened by supervisors are indeed observably different than the youth in our control group on demographics, although not in their likelihood of applying to our job posting.

3.1.2 Job Application Sample

We randomly selected 2,000 control youth and 2,000 treatment youth from our main sample to invite to apply to our job application.[18] Table A.X shows baseline balance for this subsample. Although the vast majority of baseline covariates are balanced, we note that

17. Since supervisors may not have finished the full survey, not every youth in the treatment group had a survey completed about them or a chance to get a letter. Since the names of treated youth appeared on the screen shown to supervisors after they clicked to confirm that they had been a supervisor, however, once the supervisor starts the survey, their decision about whether or not to continue with the survey may theoretically depend on the treatment status of youth. Consequently, the supervisor starting the survey is the point at which we consider the treatment youth to actually have a chance to have received our treatment (i.e., even if no supervisor ever answered the survey questions that needed to be answered for a letter to be created for them).

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.

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 below, labor market impacts for Hispanic youth are larger than for other groups. As 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 Table A.XI).

3.1.3 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 (rating category 1-4) versus high (rating category 5-7) overall ratings from supervisors (i.e., on the question about overall performance that supervisors are asked of both treatment and control youth).[19] Unlike our main sample, however, there is selection into who receives a rating based on supervisor behavior in the survey (e.g., whether the supervisor completes the survey or stops part-way through). 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; 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.

Appendix Table A.XII shows that, in fact, treatment youth are significantly less likely to have received a rating than control youth (66 versus 71 percent). 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.XIII 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

19. Low ratings are "Very Poor," " Poor," "Neutral," and "Good." High ratings are "Very Good," "Excellent," and "Exceptional."

which the supervisor rated every treatment youth and every control youth in the survey. We identify 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. Appendix Tables A.XIV and A.XV indeed show that while treatment youth are 1 percentage point less likely to appear on a completed survey overall, the observable distribution of baseline characteristics is similar both overall and within rating groups for this sub-sample.

3.1.4 Education Sample

We define two samples for the education analysis. School performance data (e.g., enrollment and attendance data) do not exist for graduates, dropouts, and students in charter or private schools. Consequently, when we measure enrollment and attendance, we focus on the 19,849 students who were enrolled in grades 8–12 in the year prior to randomization 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. This group excludes students outside of the DOE, pre-randomization dropouts and graduates, and students who temporarily stopped attending public school or had not yet joined the school district in the year before randomization. Focusing on this group limits missing outcome data to public school students who stopped attending public school, either from dropout or from an unobserved transfer out of the district. Our main analysis assigns 0 days present and 0 days enrolled to those who are missing (although we also check both outcomes conditional on having any enrollment and imputing group mean values for robustness). Among those in the education data, treatment and control youth are equally likely to appear in this "expected in high school" sample ($\beta = -0.003, p = 0.525$).[20]

The NYC DOE data also contains graduation and college enrollment information, but in a limited way. Graduation is reported only in the academic years that correspond to a student's on-time (4th), 5th, or 6th year graduation cohort; it is 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 2015–2016 and 2018–2019 (we do not have graduation or college data from 2019-20). Post-secondary enrollment is captured at a single point in time, 6 months after a student reaches what would have been their on-time graduation date, regardless of whether or not the student has actually graduated.

We do not observe which graduation cohort a student is in if they are not in our graduation or college records. To avoid selecting the sample based on an outcome, we limit analysis of graduation and college enrollment to a sub-sample of our "expected in high school" sample

20. Note that we do not yet have the information we need to calculate GPA, but we expect that we will be able to test for program effects on this outcome after some additional communication with the NYC DOE.

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 (i.e., in 4 years). This effectively excludes pre-randomization 8th graders, since they are not old enough to have graduated by the end of our data period. It also excludes pre-randomization 12th-graders from the college analysis, since their on-time graduation date makes their college outcome a baseline characteristic (measured just before our letters were distributed). Anyone else expected in high school but missing from the data receives a 0 for graduation or college enrollment. There is no treatment-control difference in the probability of appearing in either sample ($\beta = -0.005, p = 0.444$ for graduation and $\beta = -0.001, p = 0.905$ for college).

Note to NBER organizers: we note that our access to the employment data is restricted to when we are on Penn's campus, and restrictions over the summer made it difficult to spend time on campus. We therefore have not been able to integrate the education and employment data. Future analyses will assess joint education and employment outcomes to see if the same people are driving both sets of effects.

3.2 Outcomes of Interest

3.2.1 Labor Market Outcomes

We pre-specified a primary focus on annual earnings, Winsorized to deal with outliers and showing alternative methods to adjust for skewness, with an indicator for any employment as a secondary outcome. Our main analysis shows employment and earnings in the first year (4 quarters) after randomization 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.I.A we also show robustness to alternative transformations (Winsorization at the 99^{th} percentile and adding 0.1, 10, and 100 to earnings prior to logging).

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 outcomes in Gelber, Isen, and Kessler (2016)). These results are discussed in the Appendix of this paper.

3.2.2 Job Application Outcomes

For the sub-sample of individuals who 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 unpack the mechanisms behind any potential changes in labor market outcomes. The proclivity to opt into consideration for the "more selective" job (i.e., with higher pay that requires a stronger application) 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 an implied "first stage" on use of a letter in a job application. We can use this implied first stage to back out an implied LATE for the compliers who are not directly observed in our data: those whose job-seeking behavior was changed by random assignment.

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.

3.2.3 Education Outcomes

There are two reasons a letter of recommendation might affect education outcomes. The first is a direct effect: We encouraged young people to show their letters to their teachers and guidance counselors, and prior evidence suggest that adults' beliefs about students can directly change students' performance (Rosenthal and Jacobson 1968). The second is an indirect effect: If the letters encourage youth marginally attached to school to join the labor market, they may shift out of education and toward work. The effect of working while in high school on educational performance has long been of interest (e.g., Ruhm 1995; Baum and Ruhm 2016, but difficult to identify in the absence of the kind of exogenous shock to employment that our letters create. We matched our sample to data from the NYC DOE to test the effect of the letters on attendance, grades, graduation, and post-secondary enrollment.

Because we knew much less about what education data would be available at the time of pre-specification, these outcomes 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, there is not enough data availability to construct the index. We do not currently have GPA available; many students are not in school to have attendance or attend a program where DOE does not capture attendance; and the test-taking pattern leaves many students without test scores.

Instead, we show the available outcomes for sub-samples on which we can best measure them: for those enrolled in 8th–12th grade during the pre-randomization year who had not

yet graduated, we study an indicator for any enrollment and counts of days enrolled and days present; for those old enough that we could have observed their potential on-time graduation, we study indicators for on-time graduation, ever graduating during the follow-up period, and an indicator for ever graduating or still attending school; and, for the smaller group where post-secondary outcomes are observed (those observed at the time they could have graduated on time, regardless of whether they did), an indicator for any college enrollment (including 2-year and 4-year institutions but excluding vocational or public service post-secondary activities).

3.3 Analytical Method

3.3.1 Main Analysis

We conduct an intent-to-treat (ITT) analysis by regressing each outcome variable on a treatment indicator and baseline covariates. The analysis equation is of the form:

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

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, including an indicator variable for study cohort since randomization occurred separately by study year.[21] For the moment we report heteroskedasticity-robust standard errors.[22]

Not every treatment youth on a started survey actually received a letter, either because: (1) they were on a survey answered by someone who was not their direct supervisor or by

21. To reduce possible misspecification within the finite sample, we include covariates as a series of indicator variables. Covariates included in the labor market analysis include indicator variables for being male; being employed in each of the 2nd through 6th pre-randomization years or never being employed prior to randomization; having annual earnings in each of the 2nd through 4th quartiles of the distribution in the year prior to randomization; self-reporting being in high school, a graduate, or in college at the time of application; not being a U.S. citizen; being age 15-16, 17-18, 19-20, or 21 and over; being part of the Ladders for Leaders program; and self-reporting being Hispanic, Asian, White, other, or having missing race/ethnicity data. Because data access restrictions currently prevent us from combining the employment and education data, the education analysis uses a different set of covariates measured only within the education data: indicators for being male; age 13-14, 17-18, 19-20, or 21 and up; being Hispanic, Asian, White, other, or missing race/ethnicity; being enrolled in grade 9, 10, 11, or 12 in the year prior to randomization; and belonging to quartiles 2, 3, or 4 of the distribution of the share of enrolled days present pre-randomization. We will also show results without covariates as a robustness check, but all hypothesis tests in the paper adjust for study cohort.

22. Once COVID protocols on Penn's campus are relaxed, we will retrieve individual identifiers to label the small number of youth who appear more than once in the data, which is typically from being part of the study in both study years. Once we have these identifiers, we will cluster our standard errors on the individual to account for the non-independence of outcomes within person across different years. Because this happens relatively rarely, we do not expect this clustering to a substantive difference to inference.

someone who did not provide ratings for them, (2) their ratings were not good enough to receive a positive letter, or (3) they never received either the mailed or emailed letter we sent. As a result of this kind of non-compliance, the ITT will understate the effect of actually receiving a letter. In addition, we cannot observe who actually uses the recommendation letters in practice, so we cannot directly estimate the effect of letter use on compliers. 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 the first 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, and we scale the ITT effect by this estimate of letter usage. 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 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 TOT 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 (i.e., rather than a direct estimate).

3.3.2 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. The main text focuses on separating labor market impacts for White and minority youth. The appendix also reports heterogeneity by age, school enrollment/graduation (as self-reported on the application to SYEP), and gender.

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 higher ratings to translate into better labor market outcomes.

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.

4 Results

4.1 Summary Statistics and Balance

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 significant. Study participants reflect the population that participants 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 are still in high school. About 45 percent of participants never appear in the UI data prior to their participation in SYEP, but during the SYEP year, 97 percent work, earning an average of just under $2,400.[23]

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 1.2 percentage points (1.8 percent relative to the control mean of 70 percent) during the year following letter distribution.[24] Actually being sent the letter increases year 1 employment by 3.1 percentage points (4.4 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 does not disappear over the cumulative two-year follow-up: being sent a letter increases net two-year employment rates by 1.9 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 imprecise. Those sent a letter a letter earned a total of $410 more over the two-year follow-up period, a very similar proportional change to the increase in employment (4 percent), though just over traditional thresholds of

23. In theory, everyone in our data should have worked in the pre-randomization year since they were all listed as a SYEP participant. Some of the non-workers may not have matched to the UI data for reasons discussed above, and some may not have ever been paid by SYEP despite being listed as a participant in their data.

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

statistical significance (p=0.12). Other transformations of earnings are more precise, with Panel C showing a significant 18 percent increase in cumulative earnings, with 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.I shows alternative level and log transformations. The results suggest that while the magnitude of the change is somewhat sensitive to functional form, the bulk of 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 0.04 quarters (0.1 for letter recipients) more than their control counterparts, and the last column shows that conditional on working at all (i.e., for those selected into work), treatment youth find jobs sooner than controls (0.12 quarters sooner for letter recipients). Together, this pattern suggests that the letters help youth find jobs sooner, but do not merely substitute early work for later work; they keep working more than they would have otherwise. One concern about this pattern is the possibility that supervisors are 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 3 suggests this is not the case: there is no increase in the number of job spells treatment youth have, either conditional or unconditional on working at all. Appendix Section A.I.B and Table A.II provide further evidence that letters are not reducing—and in fact may be increasing—the quality of the job match.

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 youths' job search intensity or confidence, or rather increase labor demand by increasing the salience or beliefs about an applicant on the employer side. By distributing our own job posting to 4,000 treatment and control youth, we are able to generate some evidence on why the letters increased employment and earnings, in addition to approximating a first stage for actually using the letter.

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.[25] The second column shows that 8.8 percent of the control group and 8.3 percent

25. 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 consider this to be "applied" because we hired

of the treatment group applied to our job. 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, the letters 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 despite our efforts to frame the job as flexible enough to be compatible with other work. It is true that if we look at the baseline characteristics of applicants, treatment applicants had somewhat less success in the labor market prior to randomization (though demographically they were quite similar to control applicants). However, 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. So the lack of supply side behavior change does not appear to be a result of the treatment youth being more likely to be employed already. Overall, the evidence from our job application suggests that labor market improvements are coming from behavior change on the demand side, 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 central assumption if employers are really the ones changing their behavior—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 use a letter at rates 250 percent higher than the control group: 1.4 percent of all those invited to apply, and 16.5 percent 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).

people even if they did not answer all the questions and hit the final submit button. There is also no treatment-control difference in fully completing the application.

This result can also help benchmark how big the employment responses are for youth who actually use their letters. If we very make the (admittedly strong) assumptions that the difference in letter use we observe in our job application applies to the entire sample and that everyone in the sample applies for 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. If we think that it might be easier to use the letter in our application than in typical job applications, that receiving a job advertisement that references SYEP and having a screen to upload supporting material might remind treatment youth about the letter or prime them to use it more than a typical job application, or that not everyone in the sample is applying for at least one job, then we are likely overstating the number of treatment youth who used a letter relative to controls. In that case, this implied TOT would use a too-large first stage, and the actual TOT would be even larger than our calculation suggests.

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 applications. 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 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 a young person, there is still variation in how positive supervisors were in their letters that allow 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

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.

of results is relatively similar for both estimators.

The first result of note is that the ratings do seem to correspond with labor market outcomes. 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. We find that the letter-driven increase in employment during year 1 is concentrated among low-rated youth, although the difference between the groups is imprecise. This suggests that initially, the presence of the letter helped those who might otherwise struggle to find work more even if the letters were not effusively positive.

But this initial boost does not last. The low-rated group has negative but not significant point estimates in year 2, such that the net employment effect is close to 0 and the cumulative earnings point estimates 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. It appears, then, that employers may be over-estimating the signal in a letter for lower-rated youth, who then do not persist in higher labor market engagement. For the marginal youth with higher ratings, on the other hand, the letter boosts them into a higher trajectory that endures. This heterogeneity suggests that reducing information frictions for those who are likely to be highly productive employees, but who might not otherwise be noticed during the hiring process, is highly beneficial. Meanwhile, less productive young people may need more intensive investment in improving skills to reap long-term gains, even if employers do respond to the increased information in the short term.

4.4 Education Effects

Table 6 shows the ITT and IV effects of the letters of recommendations for the subset of youth described in Section 3.1.4. The first three columns show that for the students in 8th–12th grade in the year prior to randomization who have not yet graduated, there are no significant changes in enrollment, days enrolled, or the share of days enrolled actually attended during the academic year in which letters are sent to you youth. Point estimates are relatively small, with confidence intervals ruling out treatment effects of more than a 2 day (0.9 percent) decrease or 0.8 day (0.5 percent) increase in days enrolled, and ruling out more than a 0.5 percentage point (0.6 percent) decrease or 0.6 percentage point (0.8 percent) increase in the share of enrolled days attended.

The only statistically significant treatment effect we observe is a decrease in on-time (i.e., four-year) graduation rates. As described in Section 3.1.4, this treatment effect is estimated among the youth expected to remain in NYC public schools for whom we could have observed on-time graduation in our post-treatment data. The ITT is a 1.5 percentage

point decline (1.9 percent), and the effect for those who were sent a letter is a 3.6 percentage point (4 percent) decrease. These results provide new evidence that encouraging labor force participation among high school students can slow down the pace at which students reach high school graduation.[27]

The next two columns suggest that this slow down may not translate into a permanent decline in graduation. The point estimates on the outcomes of whether the youth ever graduated high school, which includes 5-year and 6-year graduation outcomes when available, and on whether they are either graduated or still attending (i.e., indicating that they have not dropped out), are less than half as big as the decline in on-time graduation. The point estimates are still large enough to be substantively important—between 1 and 2 percent decreases relative to control you—but they are imprecise enough that we also cannot rule out sizable increases of similar magnitude. The same is true for college enrollment, which is only available for a smaller subset of students we observe during their on-time graduation year.

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 do, however, slow down progress towards a diploma, likely by pulling marginal students out of school and into the labor force. We likely need more time to go by—adding additional power by allowing us to measure graduation outcomes for the younger students in our sample—before we can draw clear conclusions about whether this decline in graduation is temporary or permanent.

4.5 Heterogeneity

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

The first stage results suggest part of the difference in the ITT effects is that minorities are much more likely to be sent a letter (42 versus 30 percent for minorities and Whites respectively). This difference may have to do with differential selection into the SYEP, since the program serves only 13 percent White youth, or with differential responses from the kinds of SYEP supervisors for whom minority and White youth work. But the IV results show

27. Previous work has typically measured overall graduation and run into ceiling effects, but has not been able to measure on-time graduation, as in Baum and Ruhm (2016).

that, even conditional on being sent a letter, the point estimates for minority employment and earnings are much bigger than for white letter recipients. The TOT difference is not because White letter recipients have worse letters than minority letter recipients; White youth actually receive overall ratings that are 0.5 points higher (on the 7-point scale) than minority youth, conditional on receiving a rating. But their letters seem to have less of an effect on their labor market outcomes, either because of how they use them or how employers respond to them.

Table 8 shows a similar pattern for the education outcomes. The decline in on-time graduation is entirely concentrated among minority high school students. For minority students, there is also a marginally significant decline in school persistence—either having graduated or still being in school—after 2 years of 1.9 percentage points (2 percent) for non-White youth who were sent a letter. 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 provides some corroboration that the increase in labor market engagement is generating some decrease in school engagement.[28]

5 Discussion

The fact that 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—suggests that there are, in fact, frictions in the labor market for minority youth 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 long-term labor market benefit from the letters.

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

It is worth exploring further how employers respond to information in such letters of recommendation and whether other interventions could also mitigate information frictions in the youth labor market. In the meantime, our findings provide a key insight for those concerned about youth labor market outcomes: At least some unemployment among low-

28. Once we are able to get more time with our data, we will be able to assess whether the effects are indeed concentrated among the same youth.

income, minority youth seems to be due to frictions surrounding the availability of existing information about youth, and providing that information can improve youth employment over several years.

One concern about more widespread use of credible signals like letters of recommendation could be 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. Nevertheless, 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 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. Finally, if employers end up leaving some vacancies open in the face of too much uncertainty about applicants, a widespread information-sharing intervention might increase overall employment.

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. We find that that letters lead to a decline in on-time graduation and possibly overall graduation for minority youth. In addition to being important for any future decisions about letter usage, 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 can more convincingly estimate the educational impacts of the marginal student working than prior work using observational or quasi-experimental variation. It appears that youth do substitute work for education on the margin. 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 returns to high school graduation. It seems likely that then net effect may not be beneficial. In that case, reducing employment frictions is likely to have a higher net benefit for those who have already finished their high school careers.

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.

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

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.

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.

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.

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.

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. 1995. "The extent and consequences of high school employment." *Journal of Labor Research* 16, no. 3 (September): 293–303. Accessed February 22, 2019.

———. 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	21695	21714	
Age	17.2	17.2	0.642
Male	0.427	0.427	0.991
Black	0.409	0.411	0.805
Hispanic	0.289	0.289	0.945
Asian	0.129	0.130	0.735
White	0.124	0.125	0.756
Other Race	0.049	0.045	0.080
In High School	0.755	0.751	0.340
HS Graduate	0.044	0.042	0.203
In College	0.173	0.180	0.082
Non-citizen	0.062	0.063	0.754
Not in UI Data	0.009	0.011	0.126
Never Employed Pre-SYEP	0.450	0.457	0.114
Ever Worked, Year -4	0.153	0.149	0.211
Earnings, Year -4	318	320	0.885
Ever Worked, Year -3	0.267	0.266	0.841
Earnings, Year -3	585	585	1.000
Ever Worked, Year -2	0.437	0.435	0.628
Earnings, Year -2	1072	1050	0.410
Ever Worked, Year -1	0.966	0.966	0.797
Earnings, Year -1	2379	2368	0.720
Joint F-test for all baseline characteristics	F(23, 43384) = 1.156, p=.274		

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 heteroskedasticity-robust standard errors.

Table 2: Labor Market Effects

Year	1	2	Cumulative
Panel A: Employment			
ITT	0.0124***	0.0058	0.0077**
	(0.0041)	(0.0041)	(0.0034)
CM	0.701	0.72	0.841
Sent Letter (IV)	0.0307***	0.0144	0.0191**
	(0.0102)	(0.0102)	(0.0083)
CCM	0.698	0.728	0.841
Panel B: Earnings, Winsorized at 99.5th Percentile			
ITT	55.22	110.66	165.89
	(45.65)	(73.17)	(106.73)
CM	3579	5964	9543
Sent Letter (IV)	136.55	273.64	410.19
	(112.84)	(180.87)	(263.82)
CCM	3747	6170	9917
Panel C: Log(Earnings + 1)			
ITT	0.093***	0.059*	0.074**
	(0.033)	(0.035)	(0.030)
CM	5.61	6.08	7.33
Sent Letter (IV)	0.230***	0.145*	0.182**
	(0.081)	(0.087)	(0.073)
CCM	5.64	6.19	7.39

N = 43,409. Winsorization in Panel B recodes each quarter's earnings to the 99.5th percentile of all quarterly earnings before summing across years. 73 observations adjusted in year 1, 254 in year 2, and 269 cumulatively. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors 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	0.044**	0.018	0.002	-0.049**
	(0.022)	(0.014)	(0.014)	(0.020)
CM	3.46	1.98	2.36	2.19
Sent Letter (IV)	0.109**	0.044	0.005	-0.120**
	(0.054)	(0.034)	(0.034)	(0.048)
CCM	3.59	1.98	2.35	2.18
N	43409	43409	36647	36647

Notes: Spells defined as consecutive quarters with earnings from same employer. Time to first quarter conditional on spells > 0. CM shows control means; CCM shows control complier means. Regression includes baseline covariates. Robust standard errors 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	-0.006	-0.005	-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)	-0.015	-0.013	-0.025	0.007	0.026***
	(0.024)	(0.022)	(0.017)	(0.018)	(0.007)
CCM	0.133	0.117	0.081	0.064	0.007

Notes: N = 4,000. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

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

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Low Ratings		0.0246*	-0.015	0.001	75.75	-136.23	-60.48
		(0.0133)	(0.0133)	(0.0109)	(129.08)	(211.23)	(306.76)
ITT, High Ratings		0.0136	0.0240***	0.0173**	100.50	346.80**	447.31*
		(0.0087)	(0.0087)	(0.0070)	(99.11)	(164.47)	(235.80)
P-value, test of diff.		0.489	0.014	0.208	0.879	0.071	0.190
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	0.3304***	0.0744*	-0.0455	0.0029	229.30	-412.47	-183.17
	(0.0103)	(0.0404)	(0.0403)	(0.0330)	(390.32)	(639.17)	(927.85)
IV, High Ratings	0.8109***	0.0167	0.0296***	0.0213**	123.96	427.48**	551.44*
	(0.0057)	(0.0107)	(0.0107)	(0.0086)	(122.08)	(202.65)	(290.52)
P-value, test of diff.	0	0.167	0.072	0.589	0.797	0.210	0.450
CCM, Low		0.612	0.758	0.823	2974	5848	8822
CCM, High		0.712	0.716	0.843	3633	6065	9698

Notes: To avoid selection into who is rated within a survey, sample includes only youth who were on a survey where employer rated all listed youth (n = 13,911). Low includes rating categories 1-4; high includes 5-7. Test of difference shows p-value for null hypothesis that treatment effects are equal in low- and high-rated groups. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 6: Education Effects

	Ever Enrolled Y1	Days Enrolled Y1	Share Days Present Y1	On-time Graduation	Ever Graduated	Graduated or Still Attending	On-time College
ITT	-0.0004	-0.3233	0.0005	-0.0154***	-0.006	-0.0066	-0.0081
	(0.0031)	(0.5772)	(0.0030)	(0.0051)	(0.0047)	(0.0044)	(0.0067)
CM	0.939	148.9	0.823	0.804	0.853	0.878	0.693
N	19849	19849	19849	14811	14811	14811	14066
Sent Letter (IV)	-0.001	-0.7826	0.0011	-0.0355***	-0.0138	-0.0151	-0.0187
	(0.0076)	(1.3968)	(0.0073)	(0.0117)	(0.0109)	(0.0101)	(0.0155)
CCM	0.950	152.5	0.847	0.856	0.887	0.912	0.738
N	19849	19849	19849	14811	14811	14811	14066

Notes: Analysis 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 5- or 6-year graduation observed during the follow-up period. College enrollment 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. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 7: Labor Market Effects for Minority and White Youth

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Minority		0.0131***	0.0065	0.0088**	73.53	148.42*	221.95**
		(0.0044)	(0.0044)	(0.0036)	(48.00)	(77.18)	(112.27)
ITT, White		0.005	-0.0011	-0.0029	-70.39	-156.55	-226.94
		(0.0115)	(0.0123)	(0.0096)	(143.64)	(221.60)	(330.99)
P-value, test of diff.		0.512	0.562	0.255	0.342	0.194	0.199
CM, Minority		0.6932	0.7229	0.839	3540	5958	9498
CM, White		0.7518	0.6949	0.851	3754	5702	9457
	First Stage						
IV, Minority	0.4188***	0.0312***	0.0155	0.0209**	175.58	354.42*	530.00**
	(0.0036)	(0.0106)	(0.0105)	(0.0086)	(114.57)	(184.25)	(268.00)
IV, White	0.2973***	0.0167	-0.0037	-0.0099	-237.61	-528.34	-765.95
	(0.0088)	(0.0387)	(0.0414)	(0.0324)	(484.04)	(746.65)	(1115.33)
P-value, test of diff.	0	0.719	0.654	0.358	0.406	0.251	0.259
CCM, Minority		0.692	0.73	0.839	3663	6093	9756
CCM, White		0.752	0.711	0.863	4402	6576	10978

Notes: N = 37,653 Minority youth and 5,366 White youth, with 390 observations dropped due to missing race/ethnicity. Earnings winsorization recodes each quarter's earnings to the 99.5th percentile of all quarterly earnings before summing across years. 69 observations adjusted in year 1, 237 in year 2, and 252 cumulatively. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 8: Education Effects for Minority and White Youth

		Ever Enrolled Y1	Days Enrolled Y1	Share Days Present Y1	On-time Graduation	Ever Graduated	Graduated or Still Attending	On-time College
ITT, Minority		-0.0003	-0.2984	0.0004	-0.0173***	-0.0074	-0.0083*	-0.011
		(0.0033)	(0.6132)	(0.0032)	(0.0054)	(0.0051)	(0.0047)	(0.0071)
ITT, White		-0.0000	-0.4347	0.0026	0.0054	0.0084	0.0117	0.0161
		(0.0087)	(1.5736)	(0.0086)	(0.0133)	(0.0127)	(0.0107)	(0.0213)
P-value, test of diff.		0.982	0.936	0.814	0.112	0.245	0.087	0.228
CM, Minority		0.936	148.2	0.816	0.796	0.848	0.874	0.682
CM, White		0.970	158.0	0.905	0.890	0.906	0.930	0.822
	First Stage							
IV, Minority	0.4189***	-0.0006	-0.7123	0.001	-0.0392***	-0.0168	-0.0189*	-0.0249
	(0.0052)	(0.0079)	(1.4634)	(0.0076)	(0.0123)	(0.0115)	(0.0107)	(0.0161)
IV, White	0.3389***	-0.0001	-1.2877	0.0077	0.0159	0.0245	0.0338	0.0464
	(0.0164)	(0.0258)	(4.6835)	(0.0255)	(0.0382)	(0.0366)	(0.0309)	(0.0613)
P-value, test of diff.	0.00	0.986	0.907	0.802	0.170	0.282	0.108	0.260
CCM, Minority		0.947	151.9	0.843	0.852	0.886	0.911	0.734
CCM, White		0.982	160.1	0.906	0.894	0.899	0.917	0.785

Analysis 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 5- or 6-year graduation observed during the follow-up period. College enrollment only measured within 6 months after a student's on-time graduation date, regardless of graduation status. For those in high school sample, 18,238 minority youth and 1,553 White youth, with 58 observations dropped due to missing race/ethnicity. For graduation sample, 13,539 minority youth and 1,214 White youth, with 58 observations dropped due to missing race/ethnicity. For college sample, 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. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Figure 1: Experimental Flow Chart

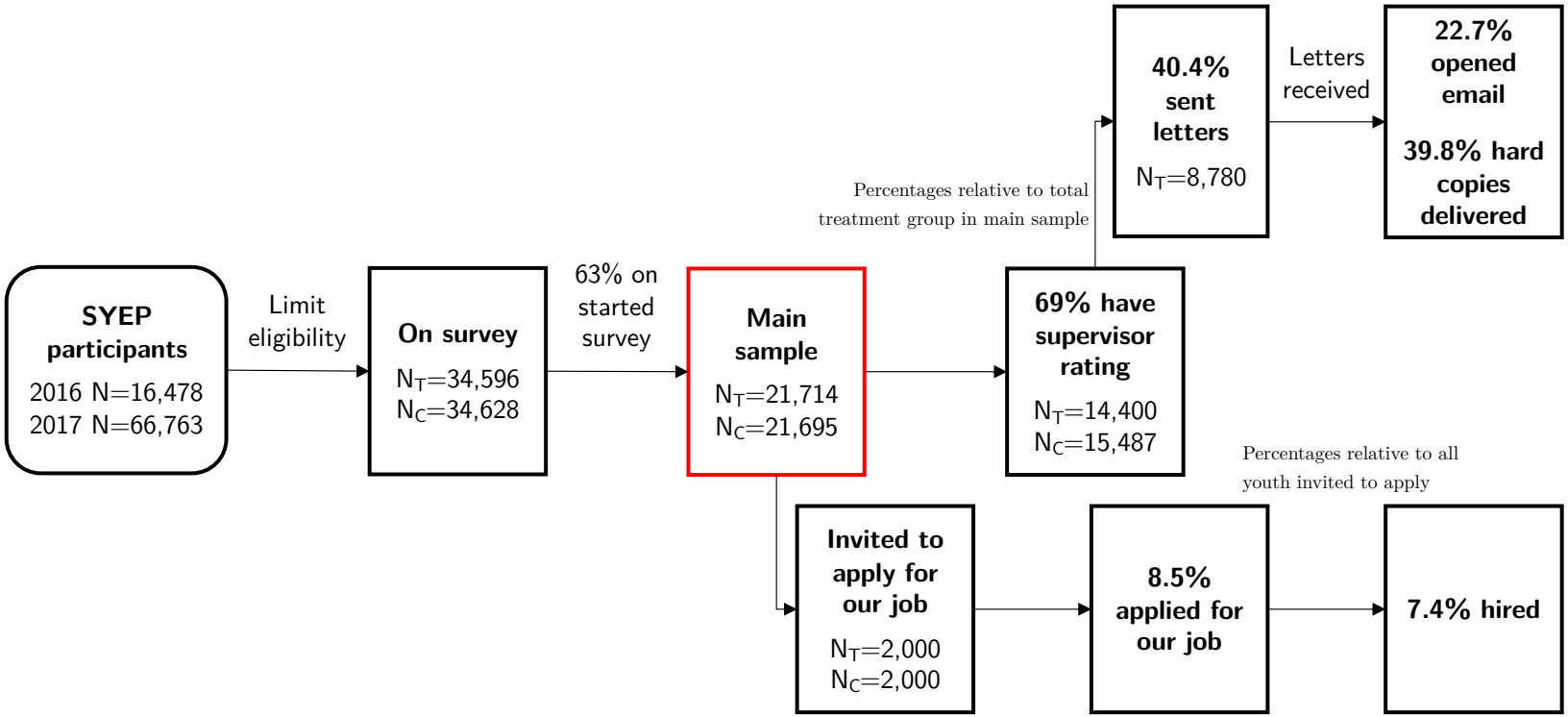


Figure 2: Screen Shots from Treatment Youth Questions on Employer Survey

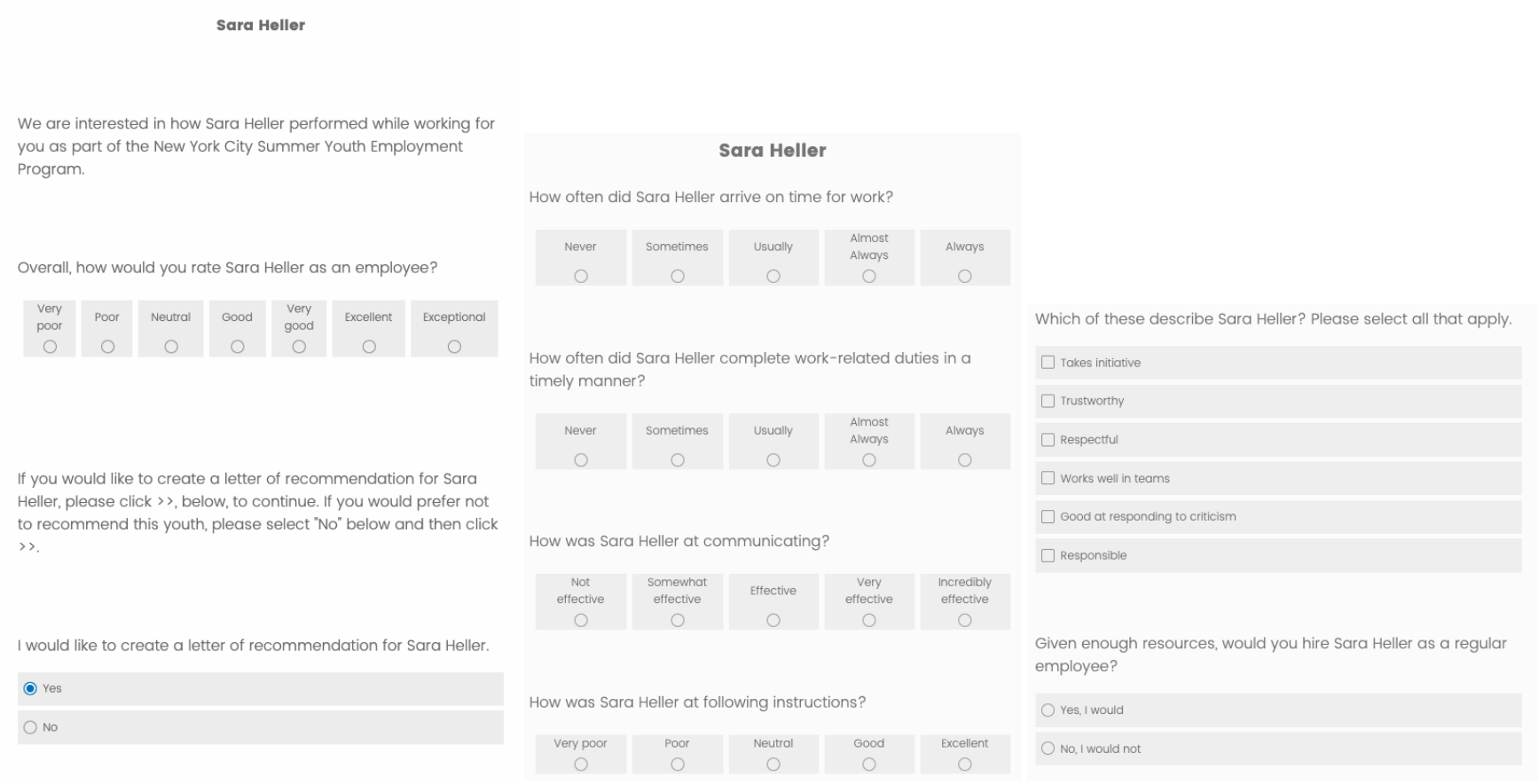


Figure 3: Example Letter of Recommendation



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.

ONLINE APPENDIX

The Effects of Letters of Recommendation in the Youth Labor Market

Sara B. Heller
University of Michigan

Judd B. Kessler
The Wharton School

A.I Additional Results, Main Sample

A.I.A Additional Labor Market Results

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.I shows other transformations of the raw dollar amounts, including an alternative winsorization (at the 99th percentile) and alternative intercepts added to earnings prior to logging (0.1, 10, and 100). 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 9.2 to 22.6 percent increases. 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.

Some additional nuance in the pattern of labor market results is shown in Figure A.VII, which breaks down both the employment and earnings results by quarter rather than by year. Although the quarterly results are fairly imprecise, several general patterns are of interest.

First, we would expect the cumulative employment effect to go to zero over time, as 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.I.B 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 over-update (i.e., believe treated applicants will be more productive workers than they actually will be). 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, allaying some concerns that over-updating could generate bad matches and increase churn.

Table A.II provides 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). 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 or about 6 percent for letter recipients. The lengthening of spells provides some suggestive evidence that letters are not generating bad matches that terminate more quickly, but rather may be improving match quality.

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. The last 3 columns of the table confirm that the results are robust to looking only at spells that are not censored, reporting 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. If anything, they appear to be generating longer job durations.

A.II Heterogeneity

A.II.A Employer Type

Tables A.III and A.IV 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.III shows this is not the main driver of the program effects 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. Cumulative employment effects are fairly similar at both types of employers, suggesting some increased return to SYEPs or other programming run by the same agency accompanied the increase in other employment. Earnings impacts are directionally much larger at non-DYCD employers, on the order of 5 rather than 0.5 percent, though too imprecise to draw strong conclusions.

Table A.IV 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 more commonly part of SYEPs like childcare and landscaping (cluster 1) separately from industries under-represented in SYEPs such as retail and food service (cluster 2). Here again, 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 not signaling that youth gained industry-specific skills, nor encouraging youth to apply to jobs more like their summer experiences. Rather, it seems the letters signal that youth are likely to be successful employees more generally.

A.II.B Race/Ethnicity, Gender, and Student Status

Tables A.V through A.IX show treatment effects for different subgroups of youth. Because of the number of hypothesis tests across these tables, and the fact that we do not have

the statistical power to differentiate between groups, 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.

The main text showed that labor market results were concentrated among non-white youth. Tables A.V through A.VII shows 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.

A similar pattern holds for women (Table A.VIII), 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 NYC sample. The fact that there are larger effects for women in both settings where priors are likely to favor or disfavor women may suggest that the effect is not simply about statistical discrimination.

Lastly, Table A.IX shows effects separately for youth who report still being in high school or not at the time of application. Employment effects are suggestively bigger for high school students, which have lower control means as well, although earnings effects are more similar across the groups.

A.III Other Results

The remainder of the appendix tables are all discussed in the main text. Tables A.X and A.XI show baseline balance (including the over-representation of Hispanic youth in the treatment group) for the randomly selected subsample to whom we sent our job advertisement. Tables A.XII and A.XIII show baseline balance separately for the subsample of study youth who received an overall rating from an employer, and breaks out the balance tests for youth receiving low versus high ratings. Tables A.XIV and A.XV further limit the subsample to youth who appeared on a survey 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. This is the subsample we use to assess how treatment effects vary by rating.

A.IV Additional Figures and Tables

Figure A.I: Example Employer 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.II: Screen Shots from Beginning of Employer Survey

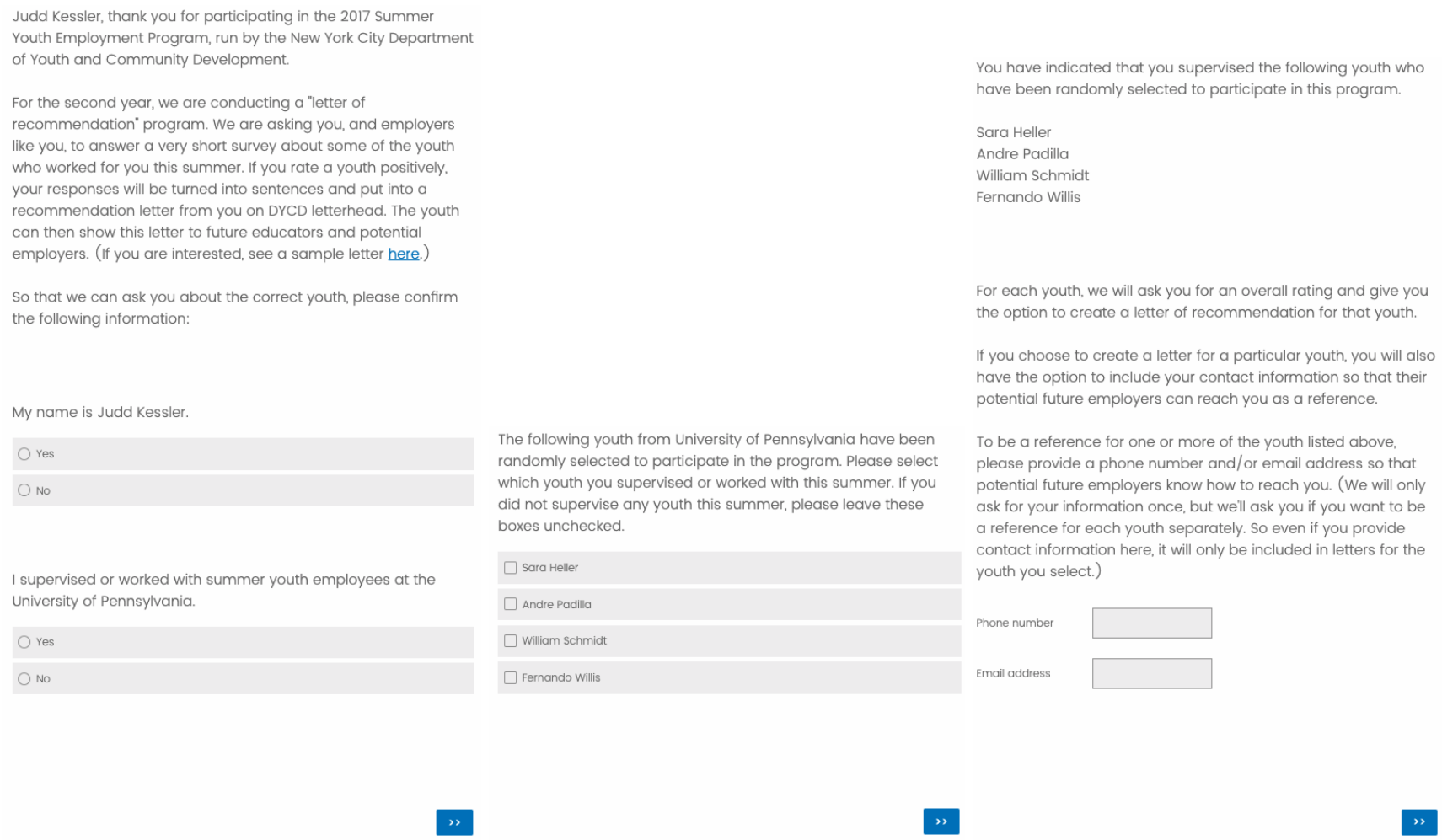
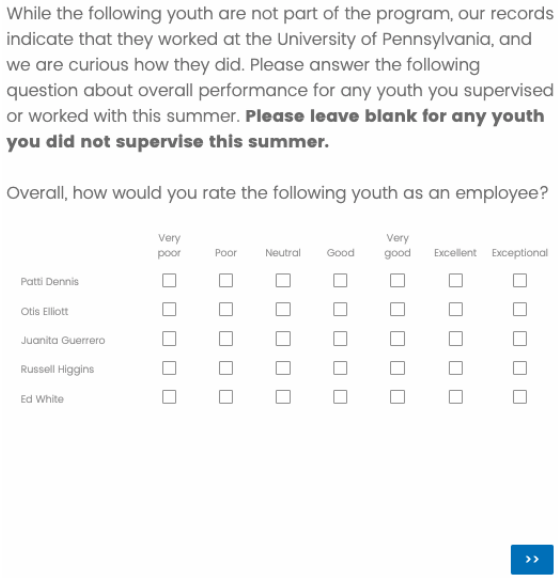


Figure A.III: Screen Shots from Control Youth Rating on Employer Survey


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.IV: Example Cover Letter to the Letter of Recommendation



November 1, 2017

Sara Heller
123 Arbor Street
New York City, NY 92323

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).

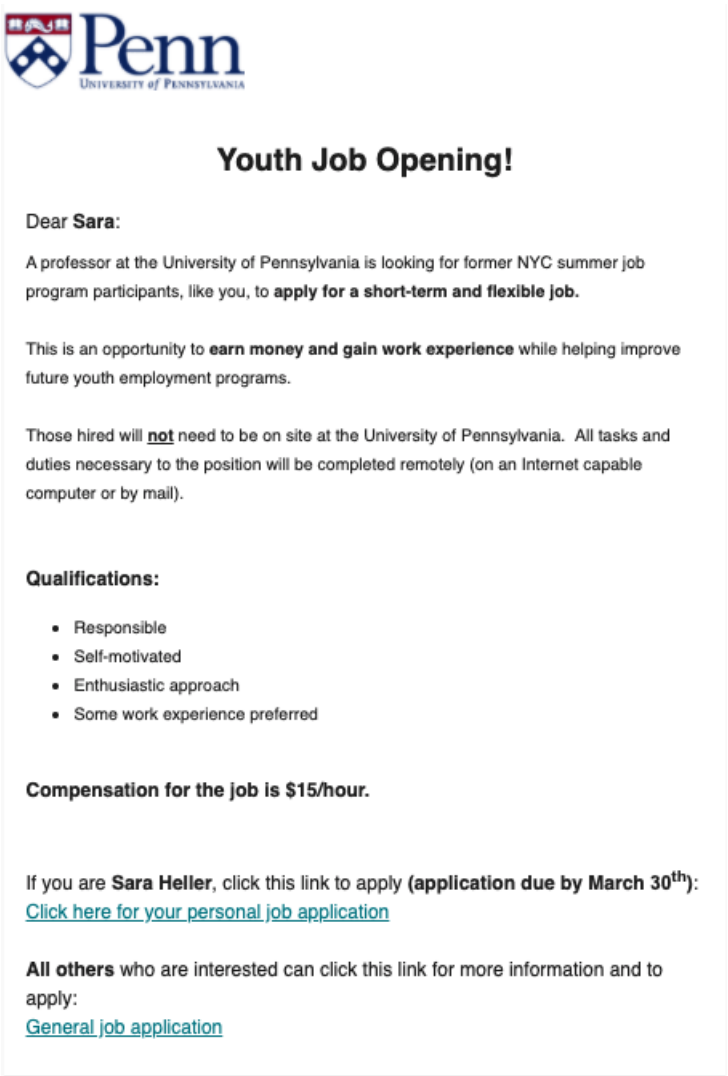
Sincerely,

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

Figure A.V: 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.VI: Job Application Prompt to Upload Supporting Documents

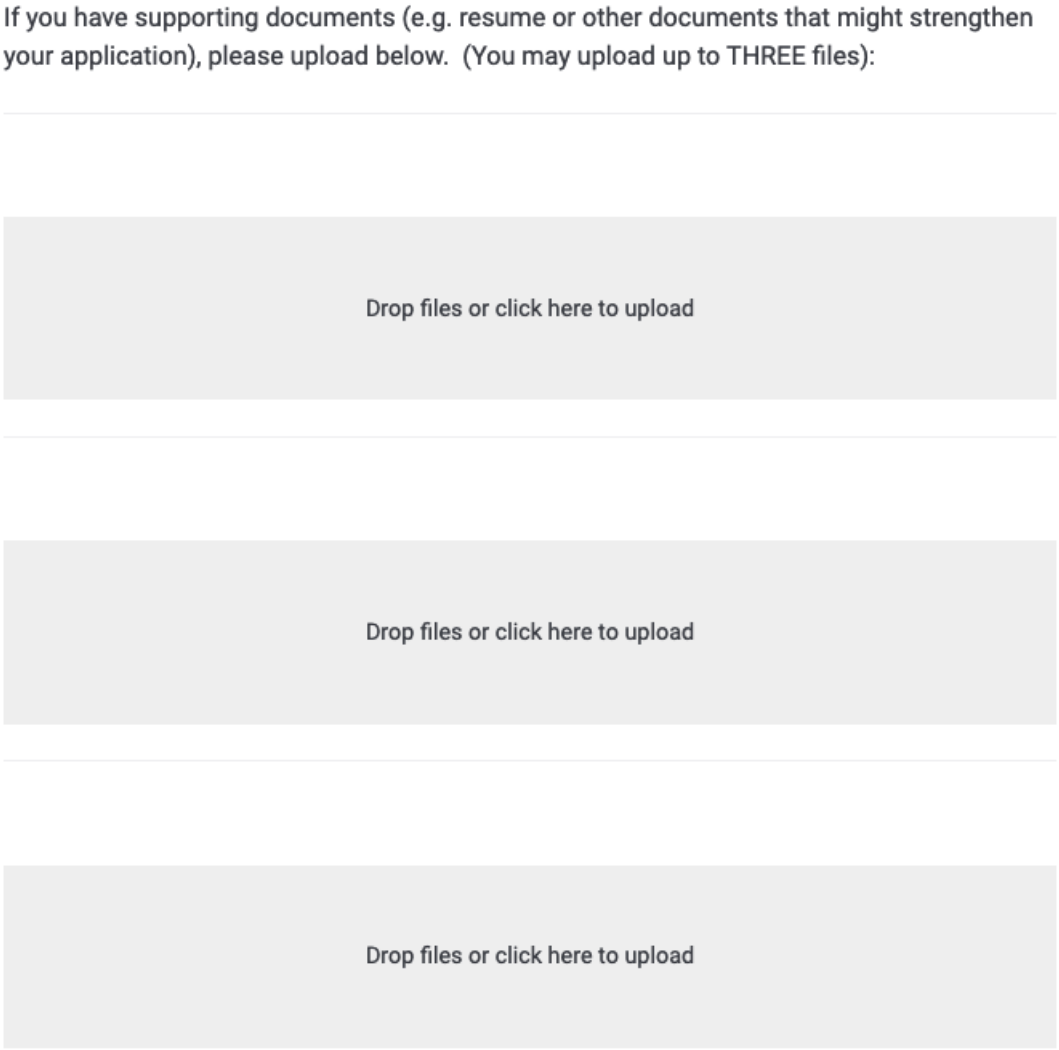
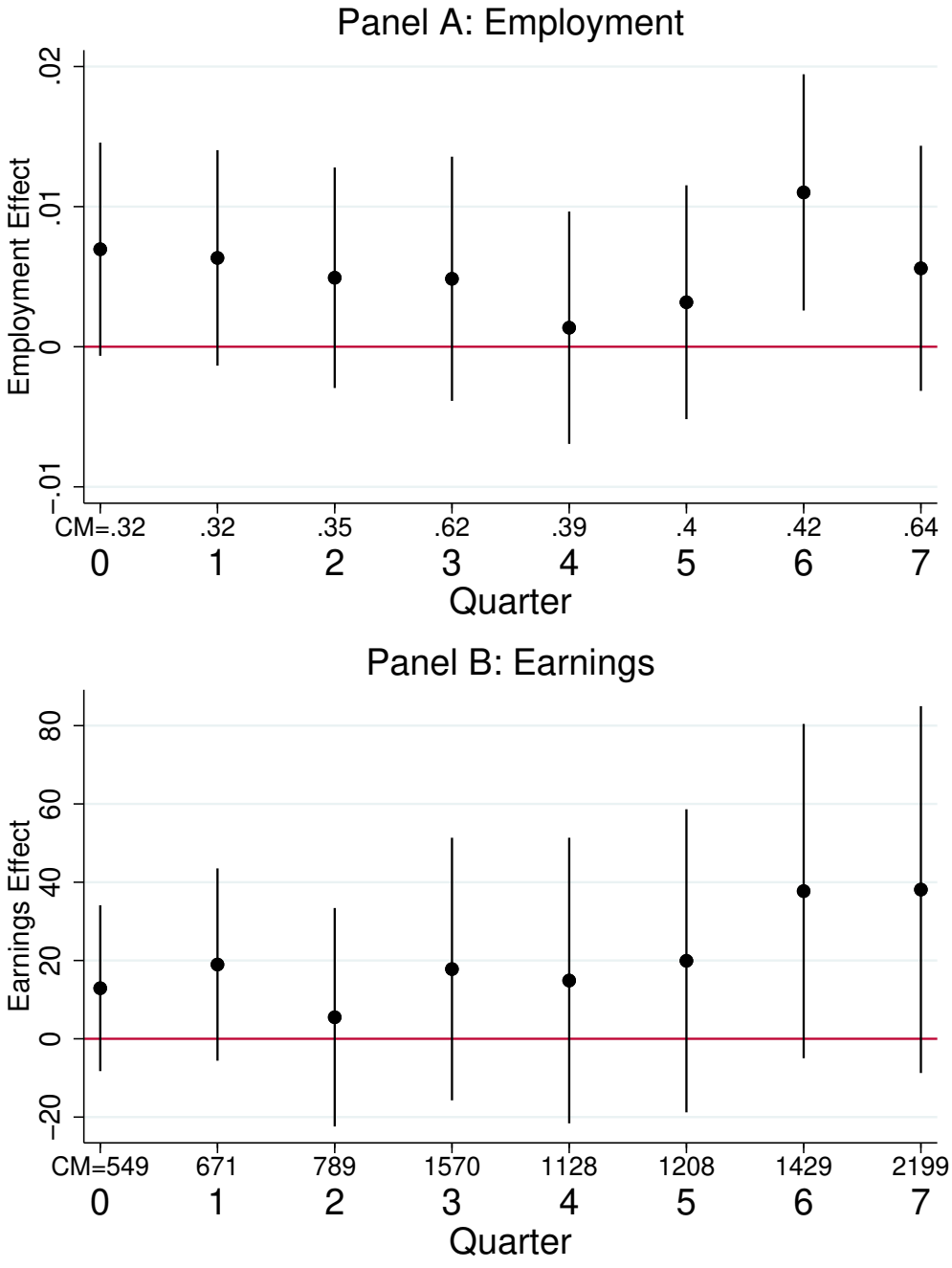


Figure A.VII: Labor Market Effects by Quarter



Note: Figures shows treatment effects on employment and earnings by quarter, with 95 percent confidence intervals calculated from heteroskedasticity-robust standard errors. CM below axis displays the control mean in that quarter. All effects from regressions that include baseline covariates.

Table A.I: Earnings Impacts across Different Skewness Adjustments

Year	1	2	Cumulative
	Panel A: Winsorized at 99th Percentile		
ITT	52.99	109.77	162.76
	(44.67)	(70.81)	(103.55)
CM	3567	5913	9479
Sent Letter (IV)	131.02	271.44	402.46
	(110.42)	(175.03)	(255.96)
CCM	3736	6121	9857
	Panel B: Log(Earnings + 0.1)		
ITT	0.121***	0.072	0.091**
	(0.042)	(0.044)	(0.037)
CM	4.92	5.44	6.97
Sent Letter (IV)	0.300***	0.178	0.226**
	(0.104)	(0.110)	(0.091)
CCM	4.95	5.56	7.03
	Panel C: Log(Earnings + 10)		
ITT	0.064***	0.045*	0.056**
	(0.024)	(0.026)	(0.022)
CM	6.30	6.73	7.7
Sent Letter (IV)	0.159***	0.112*	0.138**
	(0.058)	(0.065)	(0.055)
CCM	6.34	6.82	7.76
	Panel D: Log(Earnings + 100)		
ITT	0.037**	0.031*	0.037**
	(0.015)	(0.017)	(0.016)
CM	7.03	7.41	8.10
Sent Letter (IV)	0.091**	0.077*	0.092**
	(0.037)	(0.043)	(0.038)
CCM	7.07	7.47	8.15

Notes: N = 43,409. Winsorization in Panel A recodes each quarter's earnings to the 99th percentile of all quarterly earnings before summing across years. 159 observations adjusted in year 1, 509 in year 2, and 550 cumulatively. CM shows control means; CCM shows control complier means. Baseline covariates included in all regressions. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.II: 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	0.0219	-0.0051	0.0013	0.0062	0.0003
	(0.0290)	(0.0044)	(0.0048)	(0.0052)	(0.0050)
CM	3.22	0.23	0.63	0.47	0.37
IV	0.0533	-0.0125	0.0031	0.0150	0.0006
	(0.0707)	(0.0107)	(0.0117)	(0.0124)	(0.0122)
CCM	3.41	0.25	0.65	0.50	0.40
N	36647	36647	33546	32626	31957
			Spell 2		
ITT	0.0471**	0.0004	0.0026	0.0144*	0.0183**
	(0.0187)	(0.0059)	(0.0072)	(0.0077)	(0.0074)
CM	1.99	0.49	0.63	0.38	0.26
IV	0.1139**	0.0009	0.0063	0.0345*	0.0437**
	(0.0452)	(0.0142)	(0.0171)	(0.0184)	(0.0176)
CCM	1.98	0.50	0.64	0.39	0.25
N	25203	25203	18003	15788	14154
			Spell 3		
ITT	-0.0035	0.0062	-0.0054	-0.0060	-0.0022
	(0.0208)	(0.0085)	(0.0105)	(0.0114)	(0.0110)
CM	1.77	0.55	0.61	0.35	0.22
IV	-0.0085	0.0149	-0.0129	-0.0145	-0.0052
	(0.0496)	(0.0202)	(0.0249)	(0.0274)	(0.0261)
CCM	1.81	0.56	0.64	0.39	0.23
N	12820	12820	8660	6903	5604

Notes: Total spells conditional on having that spell. Indicators for at least X quarters conditional on observing for at least X quarters. CM shows control means; CCM shows control complier means. Baseline covariates included in all regressions. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.III: 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	0.0091**	0.0011	0.0021
	(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)	0.0114	0.008	0.0067	0.0225**	0.0026	0.0051
	(0.0114)	(0.0101)	(0.0115)	(0.0100)	(0.0106)	(0.0101)
CCM	0.419	0.253	0.507	0.428	0.588	0.638
	Panel B: Earnings, Winsorized at 99.5th Percentile					
ITT	0.47	2.98	3.45	135.27	265.87	401.13
	(10.74)	(9.97)	(16.81)	(113.73)	(182.74)	(266.81)
CM	810	572	1382	2770	5391	8161
Sent Letter (IV)	1.17	7.36	8.53	135.27	265.87	401.13
	(26.56)	(24.64)	(41.55)	(113.73)	(182.74)	(266.81)
CCM	871	574	1445	2877	5596	8473

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 earnings to the 99.5th percentile of all quarterly earnings before summing across years. 73 observations adjusted in year 1, 254 in year 2, and 269 cumulatively. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.IV: 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	0.0045	0.006	0.0024	0.0075*	0.0014	0.0009
	(0.0047)	(0.0047)	(0.0045)	(0.0040)	(0.0045)	(0.0044)
CM	0.5243	0.4407	0.6495	0.3105	0.4256	0.4869
Sent Letter (IV)	0.0112	0.0148	0.0059	0.0184*	0.0035	0.0023
	(0.0116)	(0.0116)	(0.0111)	(0.0099)	(0.0110)	(0.0109)
CCM	0.537	0.443	0.666	0.306	0.436	0.496
	Panel B: Earnings, Winsorized at 99.5th Percentile					
ITT	18.99	-0.19	18.8	32.33	112.40*	144.73
	(30.88)	(48.99)	(71.44)	(40.96)	(65.46)	(95.74)
CM	1658	2261	3919	1882	3637	5518
Sent Letter (IV)	46.95	-0.46	46.5	79.94	277.92*	357.86
	(76.34)	(121.10)	(176.59)	(101.26)	(161.85)	(236.71)
CCM	1816	2460	4275	1886	3618	5505

N = 43,409. Industry definition follows the cluster definitions in **gelberetal2016**. 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. 73 observations adjusted in year 1, 254 in year 2, and 269 cumulatively. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.V: Employment and Earnings Effects for Black and Non-Black Youth

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Black		0.0078	0.005	0.0052	23.92	98.86	122.77
		(0.0064)	(0.0063)	(0.0050)	(68.68)	(108.60)	(158.62)
ITT, Non-Black		0.0150***	0.0059	0.0088*	77.55	118.33	195.88
		(0.0054)	(0.0055)	(0.0045)	(60.93)	(98.05)	(143.13)
P-value, test of diff.		0.392	0.909	0.596	0.559	0.894	0.732
CM, Black		0.715	0.743	0.857	3575	5937	9512
CM, Non-Black		0.691	0.703	0.829	3560	5920	9480
	First Stage						
IV, Black	0.4040***	0.0194	0.0123	0.0128	59.21	244.67	303.87
	(0.0052)	(0.0157)	(0.0155)	(0.0124)	(169.91)	(268.61)	(392.37)
IV, Non-Black	0.4033***	0.0371***	0.0147	0.0218*	192.31	293.43	485.74
	(0.0043)	(0.0135)	(0.0137)	(0.0112)	(151.05)	(243.14)	(354.89)
P-value, test of diff.	0.911	0.39	0.908	0.594	0.558	0.893	0.731
CCM, Black		0.723	0.763	0.867	3744	6288	10031
CCM, Non-Black		0.68	0.704	0.824	3722	6032	9754

Notes: N = 17,636 Black youth and 25,383 non-Black youth with 390 observations dropped due to missing race/ethnicity. Earnings winsorization recodes each quarter's earnings to the 99.5th percentile of all quarterly earnings before summing across years. 69 observations adjusted in year 1, 237 in year 2, and 252 cumulatively. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.VI: Employment and Earnings Effects for Hispanic and Non-Hispanic Youth

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Hispanic		0.0155**	0.0063	0.0113*	176.81**	191.86	368.68*
		(0.0077)	(0.0077)	(0.0064)	(88.01)	(139.45)	(205.64)
ITT, Non-Hispanic		0.0106**	0.0052	0.0057	6.31	77.23	83.54
		(0.0049)	(0.0049)	(0.0040)	(53.31)	(85.59)	(124.41)
P-value, test of diff.		0.59	0.906	0.449	0.097	0.483	0.235
CM, Hispanic		0.687	0.718	0.830	3726	6365	10092
CM, Non-Hispanic		0.706	0.720	0.845	3501	5749	9250
	First Stage						
IV, Hispanic	0.4149***	0.0375**	0.0152	0.0273*	425.83**	462.22	888.06*
	(0.0062)	(0.0186)	(0.0186)	(0.0154)	(211.89)	(335.81)	(495.16)
IV, Non-Hispanic	0.3990***	0.0266**	0.0131	0.0142	15.73	193.51	209.24
	(0.0039)	(0.0123)	(0.0123)	(0.0099)	(133.59)	(214.47)	(311.73)
P-value, test of diff.	0.031	0.626	0.925	0.472	0.101	0.5	0.246
CCM, Hispanic		0.682	0.722	0.827	3787	6435	10222
CCM, Non-Hispanic		0.704	0.73	0.848	3707	6010	9717

Notes: N = 12,427 Hispanic youth and 30,592 non-Hispanic youth with 390 observations dropped due to missing race/ethnicity. Earnings winsorization recodes each quarter's earnings to the 99.5th percentile of all quarterly earnings before summing across years. 69 observations adjusted in year 1, 237 in year 2, and 252 cumulatively. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.VII: Employment and Earnings Effects for Hispanic and Non-Hispanic Youth

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Asian		0.0252**	0.0024	0.0147	-6.68	100.10	93.43
		(0.0121)	(0.0121)	(0.0101)	(117.29)	(208.46)	(289.85)
ITT, Non-Asian		0.0101**	0.006	0.0062*	64.83	111.88	176.71
		(0.0044)	(0.0044)	(0.0036)	(49.49)	(77.93)	(114.63)
P-value, test of diff.		0.24	0.782	0.429	0.574	0.958	0.789
CM, Asian		0.643	0.675	0.807	3010	5256	8266
CM, Non-Asian		0.709	0.726	0.846	3649	6026	9675
	First Stage						
IV, Asian	0.4833***	0.0521**	0.005	0.0304	-13.97	207.08	193.11
	(0.0094)	(0.0250)	(0.0251)	(0.0209)	(242.81)	(431.62)	(600.08)
IV, Non-Asian	0.3917***	0.0258**	0.0153	0.0158*	165.52	285.60	451.11
	(0.0035)	(0.0112)	(0.0113)	(0.0091)	(126.30)	(198.88)	(292.53)
P-value, test of diff.	0	0.336	0.709	0.524	0.512	0.869	0.699
CCM, Asian		0.629	0.667	0.789	3129	5156	8285
CCM, Non-Asian		0.71	0.739	0.851	3842	6318	10160

Notes: N = 5,578 Asian youth and 37,441 non-Asian youth with 390 observations dropped due to missing race/ethnicity. Earnings winsorization recodes each quarter's earnings to the 99.5th percentile of all quarterly earnings before summing across years. 69 observations adjusted in year 1, 237 in year 2, and 252 cumulatively. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

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

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, Male		0.0043	0.0097	0.0038	38.01	151.44	189.45
		(0.0066)	(0.0067)	(0.0056)	(66.06)	(108.76)	(158.16)
ITT, Female		0.0184***	0.003	0.0106***	68.05	80.27	148.32
		(0.0052)	(0.0052)	(0.0041)	(62.59)	(98.57)	(144.08)
P-value, test of diff.		0.094	0.424	0.332	0.741	0.628	0.847
CM, Male		0.658	0.659	0.802	3015	4997	8012
CM, Female		0.733	0.766	0.870	4000	6684	10684
	First Stage						
IV, Male	0.3962***	0.0109	0.0245	0.0097	95.95	382.28	478.23
	(0.0050)	(0.0166)	(0.0169)	(0.0142)	(166.69)	(274.48)	(399.14)
IV, Female	0.4106***	0.0449***	0.0072	0.0258***	165.74	195.52	361.27
	(0.0044)	(0.0128)	(0.0127)	(0.0099)	(152.37)	(239.96)	(350.75)
P-value, test of diff.	0.031	0.105	0.411	0.354	0.757	0.608	0.826
CCM, Male		0.675	0.666	0.812	3265	5211	8477
CCM, Female		0.714	0.773	0.862	4093	6859	10953

Notes: Notes: N = 18,539 male youth and 24,870 female youth. Earnings winsorization recodes each quarter's earnings to the 99.5th percentile of all quarterly earnings before summing across years. 73 observations adjusted in year 1, 254 in year 2, and 269 cumulatively. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.IX: Employment and Earnings Effects for High School and Non-High School Youth

		Employment Y1	Employment Y2	Employment Cumulative	Earnings Y1	Earnings Y2	Earnings Cumulative
ITT, High School		0.0149***	0.0107**	0.0094**	63.06	97.20	160.2518*
		(0.0050)	(0.0049)	(0.0041)	(39.40)	(66.28)	(94.30)
ITT, Non-High School		0.0049	-0.009	0.0025	31.28	151.84	183.12
		(0.0070)	(0.0075)	(0.0056)	(140.69)	(216.87)	(323.01)
P-value, test of diff.		0.242	0.028	0.315	0.828	0.81	0.946
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	0.3990***	0.0373***	0.0268**	0.0236**	158.04	243.57	401.6193*
	(0.0038)	(0.0124)	(0.0123)	(0.0102)	(98.73)	(166.08)	(236.29)
IV, Non-High School	0.4210***	0.0115	-0.0213	0.0059	74.31	360.69	435.00
	(0.0067)	(0.0166)	(0.0178)	(0.0133)	(334.10)	(515.01)	(767.05)
P-value, test of diff.	0.004	0.214	0.026	0.29	0.81	0.829	0.967
CCM, High School		0.658	0.698	0.823	2582	4567	7149
CCM, Non-High School		0.812	0.816	0.893	7079	10753	17832

Notes: N = 32,703 high school youth and 10,706 non-high school youth (with 1 observation missing education coded as non-high school). Earnings winsorization recodes each quarter's earnings to the 99.5th percentile of all quarterly earnings before summing across years. 73 observations adjusted in year 1, 254 in year 2, and 269 cumulatively. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.X: 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
Non-citizen	0.061	0.069	0.303
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
Joint F-test for all baseline characteristics	$F(23, 3976) = 1.321$, p=.14		

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

Table A.XI: Labor Market Effects, Job Application Sample

Year	1	2	Cumulative
Panel A: Employment			
ITT	0.0191	0.0153	0.0093
	(0.0127)	(0.0125)	(0.0102)
CM	0.694	0.723	0.845
Sent Letter (IV)	0.056	0.045	0.0272
	(0.0372)	(0.0367)	(0.0297)
CCM	0.66	0.708	0.846
Panel B: Earnings, Winsorized at 99.5th Percentile			
ITT	179.69	179.86	359.55
	(150.73)	(236.27)	(347.31)
CM	3664	6301	9966
Sent Letter (IV)	527.48	527.98	1055.46
	(441.03)	(691.68)	(1016.59)
CCM	3641	6220	9861
Panel C: Log(Earnings + 1)			
ITT	0.16	0.138	0.108
	(0.101)	(0.107)	(0.090)
CM	5.58	6.14	7.39
Sent Letter (IV)	0.47	0.404	0.318
	(0.296)	(0.314)	(0.263)
CCM	5.36	6.01	7.37

Notes: N = 4,000. Winsorization in Panel B recodes each quarter's earnings to the 99.5th percentile of all quarterly earnings before summing across years. 11 observations adjusted in year 1, 34 in year 2, and 36 cumulatively. CM shows control means; CCM shows control complier means. Baseline covariates included in all regressions. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.XII: Balance for Those with Employer Ratings

	Control Mean	Treatment Mean	Test of Difference
Has Rating	0.714	0.663	0.000
N	15487	14400	
Overall Rating	5.18	5.16	0.287
Age	17.22	17.19	0.347
Male	0.426	0.427	0.858
Black	0.418	0.413	0.370
Hispanic	0.287	0.289	0.683
Asian	0.131	0.137	0.106
White	0.117	0.115	0.747
Other Race	0.047	0.045	0.366
In High School	0.753	0.751	0.628
HS Graduate	0.042	0.041	0.837
In College	0.180	0.184	0.419
Non-citizen	0.064	0.065	0.856
Not in UI Data	0.005	0.005	0.901
Never Employed Pre-SYEP	0.446	0.454	0.165
Ever Worked, Year -4	0.155	0.149	0.213
Earnings, Year -4	320	326	0.787
Ever Worked, Year -3	0.268	0.268	0.991
Earnings, Year -3	582	589	0.803
Ever Worked, Year -2	0.442	0.438	0.527
Earnings, Year -2	1077	1065	0.687
Ever Worked, Year -1	0.981	0.988	0.000
Earnings, Year -1	2415	2438	0.559
Joint F-test for all baseline characteristics	F(24, 29861) = .928, p=.563		

Top row shows regression of an indicator for having an employer rating on treatment for full sample (n=43,409), adjusting for cohort with heteroskedasticity-robust standard errors. Rest of table uses subset of main sample with non-missing employer ratings. Test of difference reports the p-value from a regression of each characteristic on a treatment indicator, controlling for a cohort indicator and using heteroskedasticity-robust standard errors.

Table A.XIII: Balance for Those with Employer Ratings 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	0.086	17.25	17.25	1.000
Male	0.449	0.448	0.935	0.414	0.417	0.754
Black	0.492	0.500	0.420	0.382	0.371	0.118
Hispanic	0.292	0.294	0.837	0.284	0.287	0.678
Asian	0.099	0.091	0.225	0.147	0.159	0.015
White	0.069	0.070	0.876	0.140	0.137	0.567
Other Race	0.049	0.045	0.392	0.047	0.045	0.598
In High School	0.782	0.787	0.550	0.739	0.734	0.371
HS Graduate	0.046	0.043	0.471	0.040	0.040	0.767
In College	0.133	0.132	0.950	0.204	0.208	0.391
Non-citizen	0.058	0.055	0.439	0.067	0.069	0.492
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	0.727
Ever Worked, Year -3	0.255	0.236	0.037	0.274	0.282	0.178
Earnings, Year -3	507	477	0.432	618	642	0.486
Ever Worked, Year -2	0.424	0.403	0.037	0.450	0.454	0.524
Earnings, Year -2	995	879	0.024	1117	1153	0.389
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	0.207
Joint F-test	F(23, 9670) = 1.401, p=.096			F(23, 20169) = .76, p=.784		

Sample includes all youth with employer rating. Low includes rating categories 1-4; high includes 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 heteroskedasticity-robust standard errors.

Table A.XIV: Balance for Those with Employer Ratings, Fully Completed Surveys

	Control Mean	Treatment Mean	Test of Difference
On Complete Survey	0.325	0.316	0.066
N	7042	6869	
Overall Rating	5.22	5.20	0.497
Age	17.21	17.19	0.610
Male	0.413	0.418	0.493
Black	0.425	0.428	0.713
Hispanic	0.283	0.282	0.854
Asian	0.150	0.157	0.194
White	0.094	0.087	0.141
Other Race	0.048	0.045	0.491
In High School	0.751	0.750	0.911
HS Graduate	0.038	0.034	0.176
In College	0.188	0.193	0.479
Non-citizen	0.067	0.072	0.303
Not in UI Data	0.005	0.005	0.649
Never Employed Pre-SYEP	0.460	0.468	0.312
Ever Worked, Year -4	0.145	0.142	0.642
Earnings, Year -4	321	304	0.595
Ever Worked, Year -3	0.254	0.257	0.683
Earnings, Year -3	560	559	0.974
Ever Worked, Year -2	0.426	0.417	0.258
Earnings, Year -2	1013	993	0.648
Ever Worked, Year -1	0.983	0.988	0.029
Earnings, Year -1	2358	2304	0.317
Joint F-test for all baseline characteristics	$F(24, 13885) = .707$, p=.85		

Top row shows regression of indicator for whether an individual was on a fully completed survey on treatment for full sample (n=43,409), adjusting for cohort with heteroskedasticity-robust standard errors. Rest of table uses subset of main sample who were on a fully completed survey. Test of difference reports the p-value from a regression of each characteristic on a treatment indicator, controlling for a cohort indicator and using heteroskedasticity-robust standard errors.

Table A.XV: 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	0.455
Male	0.440	0.439	0.937	0.400	0.409	0.353
Black	0.505	0.535	0.053	0.388	0.381	0.482
Hispanic	0.277	0.258	0.179	0.286	0.292	0.492
Asian	0.117	0.111	0.572	0.165	0.178	0.091
White	0.051	0.047	0.465	0.114	0.105	0.148
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	0.809	0.037	0.031	0.143
In College	0.137	0.139	0.890	0.211	0.216	0.551
Non-citizen	0.068	0.061	0.307	0.067	0.077	0.063
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	0.223
Ever Worked, Year -4	0.134	0.125	0.393	0.150	0.149	0.963
Earnings, Year -4	268	266	0.974	346	320	0.528
Ever Worked, Year -3	0.242	0.234	0.540	0.259	0.266	0.395
Earnings, Year -3	466	479	0.838	603	594	0.854
Ever Worked, Year -2	0.407	0.395	0.436	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	0.343
Joint F-test		F(23, 4277) = .940, p=.543			F(23, 9586) = 1.021, p=.434	

Sample includes all youth on a fully completed survey. Low includes rating categories 1-4; high includes 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 heteroskedasticity-robust standard errors.

Table A.XVI: Comparing 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	0.893
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	0.013
HS Graduate	0.050	0.044	0.003
In College	0.174	0.173	0.677
Non-citizen	0.054	0.062	0.000
Not in UI Data	0.011	0.009	0.038
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	0.445
Ever Worked, Year -3	0.293	0.267	0.000
Earnings, Year -3	616	585	0.173
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
Joint F-test for all baseline characteristics	$F(23, 47483) = 32.003$, 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 heteroskedasticity-robust standard errors.

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

	Unopened Surveys	Control	Test of Difference
N	25813	21695	
Employment Y1	0.715	0.701	0.001
Employment Y2	0.715	0.720	0.255
Employment Cumulative	0.843	0.841	0.417
Earnings Y1	3665	3579	0.154
Earnings Y2	6190	5964	0.009
Earnings Cumulative	9855	9543	0.019
Joint F-test	F(5, 47501) = 5.932, 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 heteroskedasticity-robust standard errors.

Table A.XVIII: Comparing Job Application Behavior, Unopened Surveys versus Main Control Group

	Unopened Surveys	Control	Test of Difference
N	636	2000	
Clicked Link	0.090	0.104	0.294
Started Application	0.075	0.089	0.288
Uploaded Any File	0.047	0.053	0.586
Included Letter of Rec	0.003	0.004	0.745
Checked Selective Box	0.039	0.053	0.137
Joint F-Test	F(5, 2630) = .822, p=.534		

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 heteroskedasticity-robust standard errors.