NBER WORKING PAPER SERIES

INFORMATION FRICTIONS AND SKILL SIGNALING IN THE YOUTH LABOR MARKET

Sara B. Heller
Judd B. Kessler

Working Paper 29579
http://www.nber.org/papers/w29579

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021, Revised June 2022

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 data with us. Thanks to Charlie Brown, David Deming, Harry Holzer, Alicia Modestino, Mike Mueller-Smith, Alex Rees-Jones, Ana Reynoso, Basit Zafar, and the JIM group at Princeton for helpful comments. We are particularly grateful to Julia Breitman at DYCD for all of her help along the way, to Ben Cosman at DOE for all his support, and to Alex Hirsch, Ashley Litwin, and Lauren Shaw for phenomenal project management and research assistance. All views in the paper are those of the authors and do not represent the positions of any data provider or government agency. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Information Frictions and Skill Signaling in the Youth Labor Market
Sara B. Heller and Judd B. Kessler
NBER Working Paper No. 29579
December 2021, Revised June 2022
JEL No. C93,I21,J2,J48

ABSTRACT

This paper demonstrates that information frictions limit the labor market trajectories of young people in the U.S. We provide credible skill signals—recommendation letters based on supervisor feedback—to a random subset of 43,409 participants in New York City's summer jobs program. Letters increase employment the following year by 3 percentage points (4.5 percent). Earnings effects grow over 4 years to a cumulative $1,349 (4.9 percent). We find no evidence of increased job search or confidence; instead, the signals help employers better identify successful matches with high-productivity workers. But the additional work hampers on-time high school graduation, especially among low-achieving students.

Sara B. Heller
University of Michigan
Department of Economics
611 Tappan Street, Lorch Hall Room 238
Ann Arbor, MI 48109
and NBER
sbheller@umich.edu

Judd B. Kessler
The Wharton School
University of Pennsylvania
3620 Locust Walk
Philadelphia, PA 19104
and NBER
judd.kessler@wharton.upenn.edu

A appendix is available at http://www.nber.org/data-appendix/w29579

1 Introduction

The challenges that young people face in the labor market, and the bigger barriers facing Black and Hispanic youth, have been a central focus of both research and policy for over 50 years (Freeman and Wise 1982; Freeman and Holzer 1986; Heinrich and Holzer 2011; Hoynes, Miller, and Schaller 2012; Kahn 2010). Half a century of active labor market programs have spent billions of dollars trying to improve youths' labor market outcomes. Yet despite some success in U.S. sector-focused training, and more frequent success in developing countries, youth labor market programs in high-income countries quite frequently fail (Card, Kluve, and Weber 2018; Katz et al. 2020; Crépon and Van Den Berg 2016).

Theory offers one potential explanation for why it is so hard to improve labor market success among young people: information frictions may constrain youth employment, even when an applicant has the appropriate skills to succeed in a job (Altonji and Pierret 2001; Farber and Gibbons 1996; Jovanovic 1979). On the demand side, employers may have difficulty anticipating an applicant's future productivity. Combined with screening costs or the cost of hiring unqualified workers, such information frictions may leave qualified applicants unemployed, mismatched, or underpaid. In addition, if employers statistically discriminate based on age, class, or race, such information frictions could help to explain disparities among these groups. Information frictions may also be present on the supply side, if young people lack the networks, knowledge, or confidence to complete a successful job search (Gonzalez and Shi 2010; Holzer 1988).

A small set of experiments on skill signaling demonstrates that better information can improve labor market outcomes in an online marketplace (Pallais 2014; Stanton and Thomas 2016) and in the developing economies of South Africa (Abel, Burger, and Piraino 2020; Carranza et al. 2020), Ethiopia (Abebe et al. 2021), and Uganda (Bassi and Nansamba 2021). One feature of these environments is that they are relatively low-information settings where signals may be particularly scarce. On oDesk, employers have almost no verifiable information on applicants (Pallais 2014). And in the African context, high rates of youth unemployment, a high prevalence of self-employment, and the lack of clear educational signals all contribute to an environment where work and schooling histories provide relatively little information about an applicant's potential productivity or match quality (Bandiera et al. 2021; Van der Berg 2007).[1] As a result, we do not know whether information frictions

1. Donovan, Lu, and Schoellman (2020) provide evidence that it is precisely the low-information environment that explains labor market dynamics like elevated transition rates and higher turnover rates among low tenure workers in developing countries. Indeed, a more significant role for information frictions is one potential explanation for why active labor market programs seem to work better in developing countries than in the U.S. or Europe.

constrain youth employment in high-income countries, where observable characteristics such as prior work experience, attendance at a particular high school, and GPA may all convey useful signals to both employers and job-seekers.

In this paper, we provide evidence on the role of information frictions in a large U.S. youth labor market. We partner with the New York City Summer Youth Employment Program (NYC SYEP), which employs city youth to work over the summer, to run a large-scale field experiment. The intervention provided a random subset of participants with signals about their skills that they could share with potential employers: personalized letters of recommendation from their SYEP supervisors. To test whether these signals improve employment and earnings, we follow study participants for four years in administrative unemployment insurance data from the New York State Department of Labor. To identify whether changes are driven by supply-side or demand-side responses, we invite a subset of study youth to apply to a job posting and measure application rates and confidence. And to test for impacts on educational outcomes, which could be directly affected by the letters or indirectly affected through changes in labor force involvement, we follow school-aged youth for four years in data from the NYC Department of Education.

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.[2] To make letter production on this scale feasible, we invited program supervisors to complete a survey tool, developed by our research team, that automatically turned survey responses on individual participants into full-text letters of recommendation. When supervisors agreed to produce a letter of recommendation and provided high enough ratings of the youth worker, we sent that treatment youth a digital copy and five hard copies of the letter, which included, among other things, the supervisor's overall assessment of the youth worker and descriptions of relevant soft skills such as communication, reliability, and initiative.

The availability of a personalized letter of recommendation produces sizable labor market impacts. Being sent a letter increases the likelihood that a young person is employed by over 3 percentage points in the year after receiving the letter, a 4.5 percent increase relative to the 70 percent of their control group counterparts who work.[3] Employment effects fade out over the 4-year follow-up period, but not because controls completely converge: Effects on earnings grow monotonically from \$150 (4.0 percent) in year 1 to \$546 (5.3 percent) in year 4. Overall, sending a letter increases cumulative earnings over four years by \$1,349 (4.9

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

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

percent). While we cannot separate hours and wages in our data, we show that treatment youth find jobs faster, work in higher-paying industries, and have longer job spells. All these results suggest that the information in the letters increased job and match quality, even conditional on working.

That simply providing a one-page letter about an individual's skills improves employment in the short term and earnings in the long term—totalling almost $12 million in additional earnings among the treatment group—suggests that information frictions significantly hamper youths' labor market trajectories. We conduct two additional sets of analyses to understand why.

First, we assess job-seeking behavior among a subset of our sample. A few months after distributing letters to treatment youth, we invited 4,000 treatment and control participants to apply for a short-term job working for us remotely. Treatment youth were 267% more likely to submit a letter of recommendation as part of their application (4.5 percent of control applicants and 16.5 percent of treatment applicants included a letter in their application), suggesting that our letter distribution translated into substantial differences in the application packages employers actually saw. But the treatment group was 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 job opportunity. We also find that the letters did not generate observable differences in who applied, and that changes in outside labor market involvement do not appear to explain the lack of a supply-side response. The similarity in application behavior among treatment and control youth suggests that the letters work on the demand side by changing how employers view applicants, rather than increasing motivation, job search, or confidence on the supply side.

Second, to better understand the demand-side response to the letters—in particular, to assess whether employers react to the substantive signals in the letters or just the increased salience of an application when it includes a letter—we look at treatment heterogeneity by youth ability, measured by overall supervisor rating (which we collected for both treatment and control youth). Using the control group, we show that these ratings are predictive of future earnings and educational achievement, even controlling for other covariates. This pattern suggests that the information in the letter is conveying a real signal about future productivity that might otherwise be hard for employers to observe. We then show that lasting employment and earnings increases are concentrated among youth who are more highly rated by their SYEP supervisors. Providing letters to low-rated youth generates no earnings increase and only a temporary increase in employment, driven entirely by employment with the city agency that runs the SYEP. In contrast, providing letters to high-rated youth generates higher earnings for the next four years with employers outside the city's youth development agency. That employers respond to letters about high-rated youth, but not low-rated youth, underscores that they are not just reacting to the presence of a letter, but rather the informational signal contained in the letter.

Finally, we test for treatment effects on educational outcomes. Letters could have a direct educational effect if youth also face information frictions at school such that showing the letter to teachers or guidance counselors changes the way they engage with students. Prior work has shown that teachers' and other adults' beliefs about young people directly affect their outcomes, even when the information that changed those beliefs is fictitious (Rosenthal and Jacobson 1968; Bertrand and Duflo 2017). Letters could also have an indirect effect on educational outcomes, especially if working during high school crowds out time spent in school. There is a general consensus in the literature that working a small amount has a weakly positive effect on schooling, but that working more than 20 hours is harmful (Buscha et al. 2012; Vammen Lesner et al. 2022; Staff, Schulenberg, and Bachman 2010; Monahan, Lee, and Steinberg 2011; Baum and Ruhm 2016; Ruhm 1997). However, it is difficult to isolate exogenous variation in term-time employment among high school students. And prior work has often been limited to measuring final educational attainment in surveys rather than school performance or time-to-completion. This leaves open the possibility that pushing students on the margin of dropout into the labor market could harm their educational progress—at least outside of a setting that mandates continued school enrollment as a condition of receiving a term-time job (Le Barbanchon, Ubfal, and Araya, forthcoming).

For the nearly 20,000 youth in our study who we observe in New York City public high schools, we find some indication that, by pulling people into the labor market, letters of recommendation slow down—but do not appear to stop—high school graduation. The effect is clearest among the students we would expect to be most marginal, those with below-median GPAs in the pre-randomization year. That subgroup also has the biggest change on the employment margin, accompanied by declines in enrollment and GPA during the year letters were distributed. The welfare implications of pulling these youth out of school and into the labor market depend on how long earnings benefits persist and how those benefits compare to the cost associated with a longer time spent in high school (and some students are still too young to observe final graduation outcomes, so longer-term follow-up is needed).

Most broadly, our study contributes to the literature exploring how information changes the behavior of employers and workers in the labor market. In response to fictitious applications in audit studies (Agan and Starr 2018; Kaas and Manger 2012) and policy changes in the labor market (Bartik and Nelson 2019; Doleac and Hansen 2020), employers show less discrimination when they have more information about adults. Firms' use of temp agencies or other intermediaries, screening tests, and referral bonuses suggests that employers are willing to pay to elicit particular kinds of signals for certain types of job applicants (Autor 2001; Hoffman, Kahn, and Li 2018; Pallais and Sands 2016). Job seekers also adjust their search strategies in response to tailored occupational information and to their own changing reference points over time (Belot, Kircher, and Muller 2019; DellaVigna et al. 2022).

More specifically, we expand the literature on the impacts of labor market signals about worker ability. Pallais (2014) generated seminal evidence about how uncertainty generates inefficient hiring on oDesk, a problem that gave rise to oDesk's intermediary agencies (Stanton and Thomas 2016). Pallais (2014) finds that close-to-anonymous workers with no prior experience on the platform benefit from being hired and publicly rated, while those with prior work experience benefit from more detailed reviews over the next two months. We extend the idea of public performance signals to show that they also have important impacts—over at least four years—in a state-wide labor market, and in a setting where workers can endogenously choose whether to share the relevant signal.

While we are the first paper to test the impact of reducing information frictions on labor market trajectories in a developed economy—where there are many other available signals of applicant quality—our work relates to several recent studies in developing countries' labor markets. These experiments use skill certificates and performance review templates to convincingly show that ability signals shared with employers and/or job seekers can improve labor market outcomes in low-information environments (Abebe et al. 2021; Abel, Burger, and Piraino 2020; Bassi and Nansamba 2021; Carranza et al. 2020). In addition to investigating a very different labor market setting, we complement this important work in two other ways. First, rather than focusing on those who have completed their schooling and are seeking full-time work, our study includes students, which allows us to provide new experimental evidence on substitution decisions between work and school. Second, our setting allows us to use administrative earnings data rather than self-reported survey outcomes. Although this means we can not observe hours or wages, we are able to measure all formal sector employment, separately investigate results by employer and industry, and track participants in every job over a four-year period. Measuring the full labor market trajectory is crucial to identifying if the market finds other ways to learn about the productivity of control workers (and if so, how quickly). It is also central to assessing whether employers inefficiently react to the signals (e.g., if, based on their prior experience with applicants who bring letters or skill certificates, they take the information as a more positive signal than it actually is). Inefficient belief updating could result in hiring mistakes and additional churn that would be difficult to observe in point-in-time survey data.

It is important to be clear that our conclusions are about the role of information frictions in preventing young people from obtaining successful job matches. Our findings do not necessarily imply that broader distribution of recommendation letters or other signals would be welfare enhancing. In our study, supervisors decide who to rate, negative letters are disallowed, and workers decide whether to share the signal with potential employers. It seems likely that forcing broader letter distribution would generate more negative signals, which would likely have different effects than those we estimate here (i.e., our LATE is not the ATE). Beyond treatment heterogeneity, the impact of scaling up efforts to facilitate letters of recommendation—or other credible signals—will depend on general equilibrium effects that we cannot directly measure within our study. The theory literature makes clear that the welfare effects of expanding signals in general equilibrium could increase overall employment by helping employers fill vacancies they would otherwise have left open in the face of too much uncertainty, as in Pallais (2014).[4] But it is also possible that youth with recommendation letters simply displace those without them (although this is unlikely to have happened within the context of our control group, given that there are about one million 15- to 24-year-olds in the NYC labor market and we sent fewer than 9,000 letters across two years). That said, even the welfare implications of full displacement are not obvious, since policymakers may value potential distributional changes or efficiency gains from better matches, even if there were no net change in employment.

Additional research on exactly how letters change employers' decision-making and who might be displaced by the new hires would help to predict the welfare consequences of scaling up efforts to facilitate credible productivity signals. For now, this study provides new evidence on the role of information frictions in constraining young applicants' labor market success, which could limit the impact of programs designed to improve their skills and future labor market outcomes. We establish that reducing these frictions by providing credible signals that applicants can use to communicate to employers about their strengths can significantly improve labor market trajectories.

2 Setting, Experiment, and Data

2.1 Setting

We partner with the New York City Summer Youth Employment Program (NYC SYEP), implementing our experiment with youth who participated in the summer of 2016 or the summer of 2017. The NYC SYEP is administered by the NYC Department of Youth and Community Development (DYCD). Since a post-Great Recession minimum enrollment of 29,416 youth, enrollment grew steadily to nearly 70,000 youth in 2017. In our program years, the NYC SYEP provided youth with six weeks of paid work during July and August.

4. It could also encourage youth to work harder when they know letters may be forthcoming.

All NYC residents aged 14–24 were eligible to apply for the program, though 40% of eventual participants were aged 16–17. Participants in the program were provided with jobs in the private sector (45%), at non-profits (41%), and with public sector employers (14%). The NYC SYEP directly pays youth for their work with their matched employers at the New York State minimum wage ($11.00/hour in 2017). Youth payroll totaled $83 million in 2017, or roughly $1,200 per youth participant, with a total program cost of $127 million. Over 80% of this cost was funded by the City of New York, with a majority of its remaining funding coming from New York State (see *SYEP Annual Summary* 2017).

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

2.2 Letter of Recommendation Experiment

We received SYEP data from DYCD on a subset of participants from the 2016 NYC SYEP (n=16,478) and all of the participants in the 2017 NYC SYEP (n=66,763). The program data identified each youth's summer work site and the supervisor or supervisors for the youth at that work site. Using these data, we limited our sample in several ways. First, since we needed to contact supervisors to ask them to complete the letter of recommendation survey, we excluded youth supervised by someone without an email address in the data. Second, we excluded some youth at large work sites to avoid making the survey unmanageable for a single supervisor. In particular, if any supervisor was linked to more than 30 treatment youth, then we randomly selected 30 treatment youth to be included in the survey. We applied the same restriction for the control youth in the survey.[5] In total, this left a sample of 69,222 SYEP participants who were included on at least one survey. Figure 1 traces

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

through this and the subsequent steps of how youth moved through the study.

To generate recommendation letters, we built a survey tool that sent a personalized survey to each supervisor asking about the youth who they supervised that summer (i.e., the youth linked to them in the DYCD data).[6] The email inviting each supervisor to participate explained the letter of recommendation program, included a link to the personalized survey tool, and encouraged them to participate (a sample of the email from 2017 is shown as Appendix Figure A.1). Supervisors were given approximately two weeks to complete the survey, and we sent up to two reminder emails to supervisors who had not yet completed it. For the 2016 cohort, we emailed 3,297 supervisors at the end of September (initial emails went out on 09/29/16). For the 2017 cohort, we emailed 11,877 supervisors in October (initial emails went out on 10/12/17).

The survey began with a brief explanation for supervisors that if they rated a youth positively enough, their responses to the survey questions might be used to construct letters of recommendation. A link to an example letter was provided to aid in the explanation. Respondents were then asked to confirm that they had been a SYEP supervisor during the preceding summer (see screens at the start of the survey in Appendix Figure A.2). Once a respondent confirmed being a supervisor, they were shown the list of treatment youth linked to them in the DYCD data, listed alphabetically by last name.[7] Supervisors selected which youth they had directly supervised and were asked a set of questions about each selected youth in a random order. The survey asked supervisors for an overall rating of the youth's performance and whether they would be willing to answer questions that would turn into a letter of recommendation for the youth (see Figure A.2 for screenshots of the survey). If they were willing, they were also invited to include their contact information on the letter of recommendation to serve as a reference (97 percent of eventual letters included contact information). They then rated the youth on several attributes, shown in Figure 2.

After the supervisors answered questions about treatment youth, they were asked one question each about control youth—the same question about the overall rating on the youth's

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 names of others who might have supervised youth so we could include them in the letter of recommendation program as well. If more than one supervisor rated a young person, we generated the letter from the survey with the highest rating, breaking ties by prioritizing letters that included employer contact information, and then those with the most positive responses about the youth.

7. Note that confirming one's identity and position as an SYEP supervisor, prior to viewing treatment or control youth, is how we count "starting" the survey, a definition that is relevant below.

performance—all on one screen (see Appendix Figure A.3). They were told that these youth would not be included in the letter of recommendation program. A total of 5,854 supervisors (39 percent of all supervisors we emailed) opened the survey and confirmed that they had supervised SYEP youth during the preceding summer. In total, 43,409 young people were on a started survey, 29,887 (69 percent) of whom were given an overall rating by employers.

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

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

In total, we generated and sent 8,780 letters (1,805 in 2016 and 6,975 in 2017). We uploaded digital copies of these letters to Dropbox with a link sent to the youth for whom emails were known (1,737 in 2016 and 6,720 in 2017).[8] In addition, we mailed five physical copies of the letters via USPS to each youth along with a cover letter providing context and suggested uses for the letter (see Appendix Figure A.4 for a sample cover letter; similar text was sent to youth via email along with the link to the soft copy of the letter).[9] All letters of recommendation were sent in time for winter holiday hiring in the year after SYEP participation (letters were sent to youth in early-December 2016 for the 2016 cohort and in mid-November 2017 for the 2017 cohort).

8. 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 SYEP 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.

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

2.3 Job Application Data

To understand the mechanisms through which letters of recommendation might impact labor market outcomes of treatment youth, we advertised a job to a subset of the youth in our data, solicited job applications, and hired youth ourselves. We composed a job listing for a short-term, flexible, and remote paid job, 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 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 was included at the bottom of the job description (see the email invitation sent to youth with the job description in Appendix Figure A.5).

Youth who clicked the link in the email were taken to a job application that asked a few standard contact, background, and employment experience questions. We test for treatment effects on job search behavior using whether youth click the link and whether they apply for the job. Our application also provided an optional space to upload up to three "supporting documents (e.g. resume or other documents that might strengthen your application)." The application did not explicitly mention uploading letters of recommendation, but it would have been easy for youth to upload the soft copy of the letter of recommendation provided to them in our experiment (see the screenshot of this prompt in Appendix Figure A.6).[10] 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 across 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) that required a stronger application. The application made clear that being considered for the more selective position would not affect their chances at being selected for the regular job.

All those who submitted an application that included their name, email address, and at

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

least 1 additional field were hired.[11] The job itself was an online survey of multiple-choice questions. These questions asked youth about their experiences job-seeking and considering college, as well as about their career and education goals. At the end of the survey, there were free-response questions about the youth's experience in SYEP.[12] Workers were instructed to finish everything they could within a two-hour time frame. All youth who initiated the job-task (n=227) were paid for two hours of work via a mailed, pre-loaded debit card (so our job does not appear in the administrative data on employment and earnings).

2.4 SYEP Administrative Data

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

2.5 NYS Department of Labor Data

We obtained earnings and employment data from the New York State Department of Labor (NYSDOL). Data come from NYSDOL's quarterly Unemployment Insurance (UI) dataset, which covers formal sector employment, excluding self-employment or farming income. The data include employer name, employer FEIN, employer address, employer NAICS, and amount paid in each quarter. NYSDOL analysts matched SYEP participants to UI data using social security number. When multiple profiles in the NYSDOL data shared the same social security number, we used name to disambiguate the UI data. In total, 99.3 percent of SYEP youth in our letter of recommendation experiment were matched to the NYSDOL

11. 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 free-response questions in addition to fulfilling the requirements for the standard job.

12. Youth hired for the more selective job were asked additional free-response questions that required more thoughtful consideration.

data with no difference between treatment and control youth ($\beta = 0.001, p = 0.209$).[13]

We have data from Q1 (January–March) of 2010 through Q3 (July–September) of 2021. This window provides considerable baseline data as well as four years of outcome data after letters were sent to SYEP participants in our treatment group for each study cohort.[14]

2.6 NYC Department of Education Data

Education data come from the NYC Department of Education (DOE).[15] 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 2020–2021 school years, inclusive. SYEP applicants fail to match because they never appear in the DOE system (e.g., always attended private school), matched to more than one student record (DOE treats multiple matches on the same name and birth date as a non-match), or because typographical errors or name changes prevented identifying a study participant's education records. Overall, 88 percent of our sample matched to a DOE student record, with no treatment-control difference in match rates ($\beta = -0.003, p = 0.359$). Within the sample that matched to a DOE student record, 7,642 had no active enrollment within our 2015–2021 data. These students were largely old enough to have left school prior to 2015 (their average age at randomization is 19.7), although some may have transferred to private or non-NYC districts prior to the start of our data. This leaves 69.9% of our sample with at least some education information in the data, with no treatment-control difference ($\beta = -0.003, p = 0.442$).

3 Method of Analysis

This section discusses how we perform the analysis in this paper. In Section 3.1, we describe our sample definitions and our outcomes of interest for each data source. In Section 3.2, we describe our empirical approach, including our regression specifications. In all sections, we note cases where we deviated from our pre-analysis plan with accompanying explanations for these choices.[16]

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

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.

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

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

3.1 Sample Definitions and Outcomes

3.1.1 Labor Market Sample

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

We pre-specified this subsample of youth on a started survey as a key sample of interest, because neither treatment nor control youth on *unopened* surveys could have actually received treatment. This kind of non-compliance mechanically reduces statistical power and is orthogonal to treatment status, so we focus on the subsample with a first stage of 0.404 (rather than the first stage of 0.254 when we include youth on unopened surveys).[17] As a result, the treatment effect of receiving a letter of recommendation in our main sample is representative of the population of youth whose supervisors both had an up-to-date email address in the DYCD data and were willing to click on an invitation to participate in the letter of recommendation program. The estimates from this sample of youth almost certainly 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, and supervisors select into responding.[18]

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

17. While we pre-specified this subsample as a key sample of interest, our main sample included all SYEP participants that we randomized, because we did not anticipate that only 39% of supervisors would open the survey and that such a large fraction (i.e., over one-third) of the sample would be on an unopened survey. For completeness, we present and discuss results for this larger sample in Appendix Section A.6; Table A.26 shows main labor market results are quite similar, but slightly less precise. We choose to emphasize the results from our smaller sample in the main text, because the power gains from focusing on this subsample give better insight into the effect of the letter of recommendations on the sample of youth who might actually have been eligible to be treated, given the actions of their supervisors.

18. Appendix Section A.6 shows that youth who were on unopened surveys are indeed observably different than the youth in our control group of opened surveys on demographics and employment outcomes, although not in their likelihood of applying to our job posting. As such, it is plausible that forcing supervisors to rate youth would have somewhat different effects than those we estimate here.

As discussed below, Table 1 shows that our main sample is balanced across treatment and control youth.

3.1.2 Labor Market Outcomes

We pre-specified a primary focus on annual earnings, winsorized to deal with outliers, and an indicator for any employment as a secondary outcome. For robustness, we also show raw earnings.[19] Results based on alternative methods of adjusting for skewness are presented in Appendix Section A.1.1. Our main analysis shows employment and earnings in each of the 4 years after randomization and counts the quarter the letters went out—the fourth quarter of the year of program participation—as the first quarter of the year (so each year is from October 1st to the following September 30th). We also show results cumulatively across all 4 years of follow-up. Note that the COVID-19 pandemic started in year 4 for the 2016 cohort and in year 3 for the 2017 cohort.

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

3.1.3 Job Application Sample

We randomly selected 2,000 control youth and 2,000 treatment youth from our main 2017 cohort who had email addresses in the SYEP data to invite to apply to our job application.[20] This subsample is balanced on observables (joint test of treatment-control difference: p=0.219).[21]

19. To prevent too much leverage from a single outlier, the raw earnings regressions top code one observation that includes over $3 million in a single quarter to the next highest raw earnings amount in the data. The adjustment takes the yearly total for year 2 for this person from just under $3.2 million to just under $214,000.

20. 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.

21. Despite the overall balance, we note that the treatment group in this subsample is significantly more Hispanic by chance (33 percent in the treatment group versus 29 percent in the control group, p=0.01). As we show in Appendix Section A.3, labor market impacts for Hispanic youth are larger than for other groups. As a result, the point estimates for employment and earnings are somewhat larger for this sample.

3.1.4 Job Application Outcomes

For the sub-sample of individuals we randomly selected to receive our job application advertisement, we pre-specified three key outcomes: whether someone applied, whether they uploaded a letter, and whether they checked the box to apply to a more selective job as a measure of confidence. Observing whether there is a treatment-control difference in application rates helps us to test whether there is a supply-side job search response behind any potential changes in labor market outcomes. The proclivity to opt into consideration for the more selective job tests for treatment-control differences in self-efficacy and motivation or confidence in their likelihood of success on the labor market. Whether applicants uploaded a letter provides a measure of how much letter use actually changed in job applications.

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

3.1.5 Rated Youth Sample

To test whether the letters convey a substantive signal about worker productivity, rather than just making applications more salient, we report labor market impacts separately based on how supervisors rated an individual's overall performance. Employers were asked about the overall performance of both treatment and control youth on a seven-point scale. We split the sample into those with low overall ratings (categories 1–4: "Very Poor," " Poor," "Neutral," and "Good") and high overall ratings (categories 5–7: "Very Good," "Excellent," and "Exceptional").

Unlike our main sample, however, there is the potential for selection into who receives a rating based on supervisor behavior in the survey. Because the survey was designed to maximize the number of letters generated, treatment youth were listed first, along with a longer, multi-page set of questions on each youth; control youth were all listed at the end of the survey on a single page, with check boxes that allowed the supervisor to quickly answer the single overall quality question about each control youth. The different positioning and survey content for treatment and control youth could change the probability a supervisor rated a particular youth. Additionally, supervisors were told (and could decide whether) their responses would be turned into a letter for treatment youth, but not for control youth. The possibility of sending a letter may itself lead supervisors to make different decisions about whether to rate a youth or which rating to give. Because of both differences, we would not necessarily expect the distribution of treatment and control youth to be identical conditional on having a rating or receiving a particular rating.

In fact, treatment youth are significantly less likely to have received a rating than control youth (66 versus 71 percent, p<.01), and the distribution of ratings is somewhat different for treatment and control (see Appendix Figure A.9). There is some indication that this is driven in part by supervisors being more hesitant to give low ratings when a letter might be produced than when they knew it would not, as the distribution of baseline characteristics is nearly statistically different across treatment and control for youth receiving a low rating (p = 0.101, see Table A.5). Because we test for treatment effect heterogeneity across low-rating and high-rating groups to assess the signaling value of the letters, this kind of selection within a rating group could potentially bias our results.

To minimize the role of selection introduced by whether a youth is rated, when we report treatment effects by ratings, we focus on the sub-sample of youth who were on a survey in which the supervisor rated every treatment youth and every control youth in the survey. There are 13,911 youth who were on such a survey (4,301 with low ratings and 9,610 with high ratings). Since everyone is rated, these surveys leave no room for treatment-control differences in who is rated within the survey. In this group, treatment youth are only 0.8 percentage points less likely to appear on a completed survey overall (31.63 versus 32.46 percent, p=0.066), a small difference that might arise because it is easier to fully complete a survey with relatively fewer treatment youth. That said, the share of treatment youth on each survey is a function of random variables and, within both the low-rated and the high-rated youth of this sub-sample, observables are jointly balanced across treatment and control (see Appendix Table A.6). Appendix Section A.2 shows that even without this sample restriction, labor market results by rating are relatively similar when using all youth with a rating.

3.1.6 Education Sample

Because we knew much less about what education data would be available to us at the time of pre-specification, the education analysis is where we deviate most from our pre-analysis plan.[22] As reported above, about 70 percent of our sample has any active record in the DOE data during the period we observe (2015–2021). In practice, however, many of these students either graduated or left school prior to our 2016 and 2017 study years. In addition,

22. We initially expected to use an index that included days present, an indicator for graduating or still being in school, GPA, and standardized test scores when available, plus a separate outcome measuring post-secondary enrollment. In practice, many elements of this index are missing for multiple reasons. Many students are not in school to have attendance, or they attend a school (including charters) where DOE does not share records; we do not have standardized test scores in the data (except for the selected group that takes Regents exams); and DOE measures graduation and college enrollment only for particular cohorts at particular times. Consequently, instead of forcing different patterns of missing outcomes into a single index, we instead present results separately for the outcomes we have.

while charter school students do appear in DOE data as having active records, DOE does not share with outside researchers any information about school engagement, performance, or graduation for charter school students.

We wish to avoid missing data from students who had already left school, transferred, or attended charter schools. But we cannot define our sample based on whether they have schooling records during outcome years, since treatment could affect enrollment. Instead, we define our high school sample using only baseline characteristics. We identify students who were in public, non-charter schools, attending grades 8–12 in the pre-randomization year, but who had not graduated by the August prior to the academic year the study took place. This is the group we would expect to see in high school records if they progressed through high school without transferring or dropping out. These restrictions exclude 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. This education sample contains 19,714 students, with no treatment-control difference on being in this sample either overall ($\beta = -0.0003, p = 0.938$) or conditional on being matched to DOE data ($\beta = -0.002, p = 0.676$).[23]

3.1.7 Education Outcomes

We define year 1 of educational data as the academic year (September–June) during which letters were distributed. Note that our data are at the annual level, but letters went out in November or December. As a result, only about six or seven months of year 1 captures post-treatment outcomes. To measure academic performance, we report whether a student was enrolled at all, the percent of days enrolled for which a student was present, and GPA in year 1. For enrollment, we assign 0s for those with no attendance, though they may have attended school outside our data coverage. For GPA, we use non-missing data only.[24]

As time passes, study youth will leave school for one of multiple reasons (graduation, dropout, or transfer). Since treatment could affect this behavior, later measures of educational performance could be differentially missing across treatment and control youth. To avoid this issue, we focus on longer-term educational attainment measures that can be assessed even for those not in school. For academic progress, we report the number of credits

23. We note that while the joint test of treatment-control differences on baseline observables is above traditional cutoffs ($p = 0.149$), there is some chance imbalance on race and pre-treatment GPA within the education data, discussed in more detail in Appendix Section A.4.1. One benefit of the post-double-selection LASSO that we use in our main regression specifications (as discussed below) is that it adjusts for chance imbalance in a principled way.

24. Having GPA data is balanced across treatment and control, $\beta = 0.0009, p = 0.845$. In addition, since there is treatment-control balance on whether someone is in the enrollment, attendance, and GPA data, alternative imputations of missing data would not change our results.

attempted and percentage of credits earned across the 4 years of follow-up data, including 0s for anyone who graduated, dropped out, or transferred. These measures give an overall sense of how long youth stayed in public, non-charter schools, and whether they failed a higher percentage of coursework.

Given the potential for labor market involvement to crowd out educational attainment, we are perhaps most interested in high school graduation. It is important to note that graduation data are not available for everyone. Per state standards, DOE only reports graduation in the academic years that correspond to a student's on-time (4th), 5th, or 6th year graduation cohort, even if a student returns to school after their 6th year. Graduation data are missing for students who transfer to a charter school; who move out of district; fall under another exclusion, such as having an individualized education plan (IEP); or who were not in a 4th–6th year graduating cohort between fall 2015 and summer 2021.[25]

We include 3 different measures of graduation. The first is an indicator for on-time (4-year) graduation. In our education sample, everyone is old enough to have observed at least their on-time graduation. The second measure is an indicator for whether someone ever graduated at any point in our outcome data. This captures later graduation, but some cohorts are not yet old enough to have reached their 5th- or 6th-year graduation date. Appendix Figure A.7 diagrams the available graduation data by grade and study cohort; about 6 percent of students in our education sample are too young to have 5-year graduation recorded, and 25 percent are missing 6-year graduation. These students will have 0s for "ever graduated," although they may still graduate in the future. Additionally, some students may take longer than 6 years to graduate, which (per state standards) is not captured in DOE data. To include information on whether these younger and older students are still pursuing a diploma, we create a third measure of "school persistence," which is an indicator for whether someone has either graduated or is still attending school in the 2020–21 academic year.

There are 865 youth in our education sample who do not appear in the graduation data, likely because they transferred out of the district or joined a different group excluded from state graduation counts after randomization. Since these individuals did not receive a diploma from NYC DOE, we assign them zeros for graduation. DOE discharge codes suggest there is no treatment effect on whether students transfer out of the district ($\beta = 0.003$, $p = 0.260$, with a control mean of 0.032). Since we do not observe graduation outside the

25. Note that the graduating cohort in DOE data is defined by the official 9th grade cohort to which a student belongs per state standards. We do not directly observe which graduation cohort students are in if they are not in our graduation records, so our education sample is defined based on pre-randomization grade rather than official graduating cohort. This means that students who transferred to other districts during the outcome period will remain in our data; we discuss their outcome definition below.

district, the balance on transfers helps to rule out the possibility of differential mobility biasing the graduation results.

Lastly, we have a measure of college enrollment. DOE captures post-secondary enrollment data at a single point in time, 6 months after a student reaches their on-time graduation date (i.e., only on-time graduates will have non-zero college enrollment recorded in the data). This information is based on data from the National Student Clearinghouse and from the City University of New York. Because of the timing of this measure, our post-secondary enrollment analysis makes one additional limitation relative to the education sample: it also excludes all pre-randomization 12th-graders from the "college analysis" sample, since their on-time graduation date makes their college outcome a baseline characteristic (measured just before our letters were distributed). There is treatment-control balance on the probability of being in this sample ($\beta < 0.0001, p = 0.998$).[26]

We define any post-secondary enrollment as whether someone is enrolled in a 2-year or 4-year institution 6 months after what would have been their on-time graduation date. We do not count participation in vocational or public service post-secondary activities as college enrollment. As with graduation, we assign a 0 from anyone who is part of the college analysis sample but missing from the post-secondary data.

3.2 Analytical Method

3.2.1 Main Analysis

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

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

where Y_{it} is an outcome for individual i at time t, T_i is an indicator for random assignment to treatment, and X_{it-1} is a vector of covariates measured at or before the time of random assignment. As pre-specified, we use a post-double-selection LASSO to select which covariates to include in each regression (Belloni, Chernozhukov, and Hansen 2014a, 2014b; Belloni et al. 2012).[27] We always include an indicator variable for study cohort, since randomization

26. We can additionally use the information on post-secondary enrollment to assess whether differential mobility is an issue for our labor market results, since we only observe UI data within New York state. For the subset of the sample with post-secondary data available, the records capture whether someone is enrolled in an out-of-state college 6 months after their on-time graduation date. The results show no evidence of differential mobility: treatment youth are no more or less likely to leave New York State for college ($\beta = -0.002$, $p = 0.692$, with a control mean of 0.065).

27. We implement this with the Stata commands pdslasso and ivlasso (Ahrens, Hansen, and Schaffer 2020). See Appendix Section A.5 for a list of the covariates we offer the LASSO, and for results without any

occurred separately by study year. Because 1,776 individuals appear more than once in the data, we cluster our standard errors on individual as identified by SSN in the SYEP data.

Not every treatment youth on a started survey was sent a letter, either because they were on a survey answered by someone who was not their direct supervisor, the supervisor did not want to provide a letter, or the supervisor provided ratings that were not positive enough for a letter to be sent. As a result of this kind of non-compliance, the ITT will understate the effect of being sent a letter. We also use random assignment as an instrument for whether a youth was sent a letter. Since we perfectly observe whether every youth was sent a researcher-generated letter, we can estimate this treatment-on-the-treated effect for everyone. We report control complier means as a baseline measure to assess proportional changes for compliers (Kling, Liebman, and Katz 2007). Below, we also provide a rough benchmark for the magnitude of actually using the letter, not just being sent a letter, leveraging data from our job application.

4 Results

4.1 Summary Statistics

Table 1 shows average pre-randomization characteristics for the treatment and control groups. No more differences are statistically different from 0 than would be expected by chance, nor are the characteristics jointly statistically different ($p = 0.201$). Study participants reflect the population that participates in NYC's SYEP. On average, they are just over 17 years old, about 43 percent male, largely identify as non-White (only 12.5 percent list being White on their application), and 75 percent report being in high school in the spring prior to the SYEP. About 45 percent of participants did not work prior to their participation in SYEP, but 97 percent work during the SYEP year, earning an average of just over \$2,300 in that year, including their earnings from SYEP.

4.2 Labor Market Effects

Table 2 reports the main labor market effects. Panel A shows that being assigned to the treatment group increases employment rates by 1.3 percentage points (1.8 percent relative to the control mean of 70 percent) during the year following letter distribution. Actually being sent the letter increases employment in year 1 by 3.2 percentage points (4.5 percent relative to the control complier mean). The point estimates in the second year after letter distribution are about half as big and not statistically significant, and they continue to shrink over time.

covariates or with all covariates as robustness checks.

The fade out on this binary outcome is perhaps not surprising, since almost all control youth will eventually work in the formal labor market at least once (indeed, about 92 percent of controls work within the 4-year period).[28] But it is important to differentiate between two possible explanations for the fade out. The first possibility is that, as is common among active labor market interventions, the control group may completely converge with the treatment group. If this were the case, it would imply that while the signals in the letters speed up the process of information sharing, comparable information about control group workers becomes available rather quickly, such that all treatment effects are short-lived. The second possibility is that the information in the letters remains valuable over time, perhaps by helping improve the quality—or match quality—of initial jobs, which could have lasting effects (Kahn 2010; Neumark 2002; Oreopoulos, Wachter, and Heisz 2012). In this case, while the control group might catch up to the treatment group on the employment margin, the signals contained in the letters set treatment youth on a better trajectory, generating lasting earnings effects.

Panels B and C of Table 2 provide initial evidence that the letters set youth on a better labor market trajectory. They show annual raw earnings (Panel B) and winsorized earnings (Panel C). In the text, we focus on the latter, since it was our primary pre-specified outcome. Although the earnings effects for the early years are a bit noisy, they are substantively large and grow monotonically in both levels and in proportional terms over time. In year 1, the point estimate for being sent a letter is about \$150 (4 percent relative to the control complier mean, $p = 0.162$), which grows to \$546 (5.3 percent, $p = 0.085$) by year 4. Cumulatively across all 4 years, those sent a letter earned \$1,349 more than their control counterparts, a 4.9 percent increase ($p = 0.049$). This effect is not just driven by an increase on the extensive margin; conditional on working, cumulative earnings still increase (in untabulated results the IV estimate $= \$1,362, p = 0.058, N = 40,088$). The time pattern suggests that information frictions generate sustained harm to labor market outcomes.

Although we do not observe hours or wages in the UI records, we can use other aspects of the data to explore what is driving the earnings increase. Figure 4 shows effects by quarter and makes clear that the increase in work is not limited to summer jobs. Summers are quarters 3, 7, 11, and 15 in the figure, and results in those quarters do not look noticeably different from the other quarters. Table 3 reports on other measures of work intensity during this period. The first column suggests that those sent a letter work an additional 0.15 quarters over 4 years (a 2 percent increase), although the result is not statistically significant ($p = 0.129$). The second column shows that there is no increase in the number of

28. It is also possible some of the fade out has to do with the difficulty of finding work at the beginning of the COVID-19 pandemic, which is reflected in the downward shift of control means in year 3 at a time when more youth should have been joining the labor force as they age.

job spells. The point estimate on the number of jobs (including 0s) is positive but far from statistically significant, partly reflecting the change at the extensive margin of working at all. Conditional on working at all (the third column), the point estimate shrinks by about half to a 0.7 percent increase. The fifth column shows that, conditional on working, treatment youth find jobs sooner than controls (0.25 quarters sooner for letter recipients). Together, these results suggest that better signals help youth shorten the job search process but do not simply substitute early work for later work or generate more churn.

The fourth column documents an increase in average spell length, which suggests that the additional work is driven by successful job matches. We measure spell length by averaging across the first 3 (non-missing) job spells after randomization. We limit our attention to 3 spells in part because the average number of spells is just over 3, and in part because after 3 spells, censoring from the end of our data becomes a larger issue.[29] Average spell length significantly increases by about 0.12 quarters (3.6 percent). The longer job spells suggest that the recommendation letters increased worker and/or employer satisfaction with the job match. This finding argues against the hypothesis that employers are inefficiently updating off of the signals in the letter and instead suggests the signal in the letter is generating helpful sorting.[30]

Appendix Section A.1.3 provides some additional evidence on job type, suggesting that the signals in the letters help youth secure better jobs. Table A.3 shows that treatment youth do not just return to the agency that runs the SYEP; employment and earnings increases are concentrated among non-DYCD employers. Table A.4 uses the industry groupings from Gelber, Isen, and Kessler (2016) to show that letters seem to shift young workers into the higher-paying industries that are typically under-represented among the jobs that the NYC SYEP offers. Overall, it appears that the signals generate a long-term increase in earnings by helping young workers find better jobs sooner and stay in these jobs longer.

4.3 Mechanisms

4.3.1 Assessing Changes in Labor Supply

A key question about the observed increase in labor market success among treatment youth is whether the letters increase labor supply by increasing youth job search intensity or confidence, or whether the letters increase labor demand by changing beliefs about applicants

29. Across all 3 spells, there is no differential censoring across treatment groups; the treatment effect on the number of censored spells is -0.004 ($p = 0.340$). Appendix Section A.1.2 shows additional details about spell length and censoring among the first 3 spells.

30. If employers' prior experience with recommendation letters led them to believe that only very high-productivity workers have such signals, they might have been induced into hiring mistakes that could have generated additional churn. In practice, the longer spells suggest the reverse.

with letters or increasing the salience of those applicants among employers. By distributing our own job posting to 4,000 treatment and control youth, we are able to generate some evidence on why the letters increase employment and earnings.

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.[31] The second column shows that 8.8 percent of the control group and 8.2 percent of the treatment group applied to our job, a difference that is not statistically significant. 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 third column of Table 4, which, adjusting for application rates, translates into 60 percent of control applicants and 51 percent of treatment applicants checking the box to apply for the more selective job).

Of course, it is possible that even though the letters did not change the rate at which young people applied to our job, they could have changed the composition of who applied. This might be the case if letters help young job seekers better target their job applications to appropriate opportunities, or if the treatment group's increased formal labor market involvement reduced their interest in our short-term, online job—despite our framing the job as flexible enough to be compatible with other work.

To assess this possibility, we test for compositional differences between treatment and control applicants on observables, and we find no clear evidence that observables are jointly correlated with treatment.[32] We also test for differences in job application behavior only among those not employed elsewhere during the quarter the job application was distributed (despite being a selected group). Even among this group, there is still no statistically significant difference in application rates or in our confidence measure ($\beta = -0.01$, $p = 0.354$ for applying and $\beta = -0.01$, $p = 0.133$ for checking the selective box).[33] We conclude that the lack of an increase in supply-side job-seeking behavior does not appear to be due to treatment changing the composition of applicants or increasing other employment. Overall, the

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

32. We test for differences between the treatment and control individuals who applied for our job by interacting each baseline covariate with an indicator for whether the individual applied, regressing treatment on all covariates and these interactions, and then testing the hypothesis that all interaction coefficients are jointly 0. The p-value of this test is 0.15.

33. The same is true conditional on being employed in that quarter: $\beta = 0.0008$, $p = 0.959$ for applying and $\beta = -0.01$, $p = 0.481$ for confidence.

evidence from our job application suggests that labor market improvements are coming from employers responding to letters of recommendation, not from changes in youth application behavior or confidence.[34]

As an important check on whether treatment youth actually use the letters we send them—a necessary condition for employers to be able to respond to the letters—the final two columns of Table 4 show treatment effects on the files job applicants uploaded in their application to our job posting. There is no detectable change by treatment in the probability that youth upload some form of supporting material. But there is a dramatic change in whether youth upload a letter of recommendation. Only 0.4 percent of the control group submits a letter, including zeros for those who do not apply (conditional on applying, this translates to 4.5 percent of control applicants submitting a non-intervention letter with their application). Treatment youth are two and a half times more likely to submit a letter of recommendation than the control group: 1.4 percent of all those invited to apply submit a letter (16.5 percent conditional on applying). Since about 40 percent of treatment youth actually received a letter, this implies that about 41 percent of letter recipients use them when they apply to a job (16.5 percent relative to 40 percent).

Given the lack of a supply-side response, it is possible that letters *only* work when employers see them. If so, we could use the observed rates at which letters are used in our job application as an implied first stage of letter use, providing a back-of-the-envelope extrapolation of how big employment responses would be for youth who actually use their letters. If we make the quite strong assumptions that the difference in letter use we observe in our job application applies to the entire sample, that treatment and control youth apply to jobs at the same rate, and that everyone applies to at least one job, then the implied first stage for letter use is a 12 percentage point increase (4.5 versus 16.5 percent among applicants). Scaling our main ITT effects by this first stage would in turn imply that the employment increase for those who use the letter is about 11 percent relative to baseline in the first year, with an additional $4,500 in earnings over 4 years. Of course, many of the assumptions involved in this benchmark could fail, including the exclusion restriction. The calculation is just intended to give a rough sense of how big employer responses would be for compliers in this simple case.

34. This is one key difference between our results and those in developing countries, which typically find supply-side responses to employer feedback or skill certifications (Abel, Burger, and Piraino 2020; Bassi and Nansamba 2021; Carranza et al. 2020). It may be that the higher level of unemployment in African countries (Bandiera et al. 2021) generates more discouragement that performance information can reverse, as in the Gonzalez and Shi (2010) model.

4.3.2 Assessing Changes in Labor Demand

The evidence so far suggests that employers are the ones responding to the signals in the letters that treated youth include in their job applications. The way that employers respond to these productivity signals is consistent with a range of models that show how employer uncertainty about applicant ability can generate inefficient hiring. In these models, young or novice applicants may remain unemployed, badly matched, or paid less than their marginal product, because only those applicants whose expected productivity exceeds an employer's cost threshold are hired or efficiently paid, where employer costs can come from risk aversion, screening costs, or the transaction costs of bad hires (Altonji and Pierret 2001; Farber and Gibbons 1996; Kahn and Lange 2014; Pallais 2014).[35] Typically in these models, unemployed but high-ability workers at the margin of being hired would like to invest in signals of their ability, which would improve market efficiency. But such workers cannot generate credible signals on their own. The existing market under-provides these signals because the employers who bear the cost of producing them do not fully internalize the gains to other employers and workers (Becker 1964), or because, in practice, novice workers under-estimate the benefits of asking for them (as in Abel, Burger, and Piraino (2020)).

Given the lack of a supply-side response, the lasting earnings effects, and the longer job spells, it seems plausible that—consistent with these models—employers are using the letters to identify high-productivity workers who they might not have hired absent the signals in the letters (e.g., because of too much uncertainty about the workers' abilities). Because our letters vary in how strong a positive signal they communicate, they can also facilitate a more direct test for whether employers' inability to identify qualified applicants—including otherwise hard-to-observe characteristics like enthusiasm and reliability—actually hinders young people's labor market trajectories.

To test the signaling mechanism, we proxy for variation in worker ability with supervisors' overall quality ratings of each youth, which we observe for both treatment and control groups. Employers can observe this variation in signal strength for the treatment group, since higher ratings correspond with a stronger introductory sentence in the letter, as well as longer letters mentioning more positive attributes. If employers are using the information in letters to update their beliefs about each individual, we would expect higher-rated youth to benefit more than lower-rated youth from receiving a letter, since the letters of higher-rated youth allow them to signal their higher ability.[36]

35. We focus on employer uncertainty here, because we do not find a supply-side response. But models incorporating uncertainty among job-seekers produce similarly inefficient hiring and match productivity (e.g., Gonzalez and Shi 2010; Mortensen and Pissarides 1999).

36. Appendix Section A.3 discusses other types of heterogeneity and their interpretations.

As a sanity check on whether ratings and letter quality could plausibly provide an accurate signal about unobserved future worker ability, Table 5 shows the correlation between supervisor ratings and future earnings within the control group, as well as the correlation of supervisor ratings with future GPA and school persistence in the education sample. Ratings are significantly correlated with all three outcomes unconditionally. And while the magnitude of the relationship gets smaller when controlling for all the other covariates we use in our regressions (including demographics, employment history, and, for the educational outcomes, prior school performance), ratings still significantly predict all these future outcomes conditional on observables. In terms of magnitude, all else equal, a one standard deviation increase in supervisor rating (1.5 additional points on the 1–7 scale) corresponds to a substantial shift in earnings and a moderate shift in education outcomes: $1,458 more in cumulative earnings, 0.45 additional GPA points, and a 0.75 percentage point increase in the probability of having graduated or continuing to attend school by the end of the data. So it appears that supervisor ratings communicate real information about expected worker productivity in their letters that might otherwise be difficult for potential employers to observe.

Table 6 shows that the impact of treatment on future labor market outcomes is larger for youth with more positive signals. The table separately estimates employment and earnings effects 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").[37] Highly rated youth were much more likely to receive a letter (81 percent versus 33 percent). So the ITT differences between the groups reflect both differences in letter receipt and differences in outcomes conditional on being sent a letter, although the substantive pattern of results is quite similar for both the ITT and TOT.

We find that the low-rated group has net employment effects close to 0 and cumulative earnings point estimates that are negative but with large standard errors. They do have a marginally statistically significant increase in employment in year 1. While this might be a chance finding given the number of hypothesis tests in the table, deeper exploration shows that this is driven entirely by increased employment at DYCD, the agency that runs the SYEP and other year-round workforce development programs (the year 1 employment impact at DYCD is a marginally statistically significant 0.027, which is comparable to the estimate reported here, which averages across both DYCD and non-DYCD employers, of

37. 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. Such letters could still be produced, however, as long as enough other attributes were rated positively.

0.0247). Given prior evidence that SYEP participation itself may lower future earnings by encouraging youth to work in the lower-paying industries that are over-represented in the summer program (Gelber, Isen, and Kessler 2016), the additional connection to DYCD jobs may help to explain the directionally negative earnings estimates. Overall, it appears letters that provided only a weakly positive signal did not change employer beliefs enough to help young people get or keep jobs outside of the government agency designed to support them.

In contrast, the high-rated group has lasting positive and significant employment effects, including in years 2 and 3, as well as consistently positive earnings effects that grow over time. Although the test of the difference between groups varies in its level of statistical significance across outcomes and time periods, enough of the treatment effects are statistically different between the low-rated and high-rated groups to have some confidence in the result that letters generate larger and more persistent labor market improvements for the high-rated than the low-rated youth.

One might wonder whether effects are smaller among the low-rated group simply because they choose not to use letters in their job applications. Results from our job application suggest otherwise (see Appendix Section A.2). For every 100 letters sent to high-rated treatment youth, we received 3 job applications that included letters. For every 100 letters sent to low-rated treatment youth, we received 4 applications including letters. This pattern suggests that low-rated letter recipients are not less likely to use letters when applying for jobs.[38]

It appears, then, that employers are receiving letters from both high and low rated youth and are using the substance of the letters as a signal about who is likely to be a productive employee (i.e., not just taking more notice of all applications that include letters). This result suggests that information frictions may be holding back relatively high-performing youth workers (relative to other SYEP participants) and that simple credible signals can help improve the labor market prospects of these youth. Meanwhile, the group of young people who did not impress their SYEP supervisors as much may need more intensive investments—such as improvements in their human capital—to improve their labor market outcomes.

A natural question, given that our entire sample is composed of SYEP participants, is about the external validity of our results. Are we documenting a general phenomenon about information in the youth labor market in New York City, or are we documenting that letters help overcome a particular stigma associated with SYEP participation? The answer to this question rests in part on whether the employers in our data know that youth applicants are SYEP participants, which is necessary for the stigma story. While we do not observe

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

that directly, we can take a hint from the applications that youth submitted in response to our job advertisement. In those applications, only 22 percent of applicants self-identify as a SYEP participant in either their list of work experience or their résumé. Given that almost 80 percent of job applicants may not appear as prior SYEP participants to employers, it seems plausible that the frictions we document are not specific to SYEP-related beliefs among employers.[39]

4.4 Education Outcomes

Unlike the existing experimental literature on labor market signaling in developing countries, which focuses on those seeking full-time employment, SYEPs serve many young people who are still in high school. The inclusion of school-aged youth in our sample allows us to explore two ways in which the letters could affect educational outcomes. First, given evidence that teachers' beliefs about students shape educational performance, even when beliefs are based on fictional information (Papageorge, Gershenson, and Kang 2020; Rosenthal and Jacobson 1968), letters could improve educational outcomes if teachers update their beliefs about a student after seeing the signal in the letter. The instructions we sent young people mentioned showing letters to teachers and guidance counselors for this reason.[40]

Second, pulling students into the labor market could pull them out of school. This kind of substitution has long been a concern in the literature on working during school. The consensus from natural variation in term-time work suggests that working fewer than 20 hours a week during school has weakly positive effects on education (Buscha et al. 2012; Staff, Schulenberg, and Bachman 2010; Lesner et al. 2022; Monahan, Lee, and Steinberg 2011; Baum and Ruhm 2016; Ruhm 1997). This consensus has led to a policy push to offer year-round employment opportunities to students. But analyzing students who choose to work could be different than pushing students at the margin (e.g., those who would not find work without the letter) into the labor market. To our knowledge, the only causal evidence on this question comes from a program in Uruguay that conditioned participation in a work program on staying in school (Le Barbanchon, Ubfal, and Araya, forthcoming), so it cannot speak to what youth would choose in the absence of a structured incentive to maintain school attendance.

Table 7 shows treatment effects for the 45 percent of our sample we expect to observe in high school outcome data (see Section 3.1.6 for the detailed sample definition). Overall, point

39. It is also worth noting that since the recommendation letters we distribute come on letterhead from DYCD—the agency that runs the SYEP—any stigma against SYEP or DYCD among employers would push against us finding a positive impact of the letters on employment outcomes.

40. A related mechanism would be if the signal in the letter changed the youth's beliefs about their own ability to succeed in school (e.g., if it boosted their academic confidence).

estimates suggest some declines in performance and attainment associated with the letters of recommendation, but they are not statistically significant. The one marginally significant result is a decline in on-time graduation (a 1.6 percentage point, or 2 percent, decline for letter recipients). The effects on ever graduating and school persistence (graduating or still attending) are considerably smaller, hinting that letters might slow down, but not stop, high school completion.

On one hand, we might not want to make too much of this result given that there is only one marginally significant change across 9 different outcomes. On the other hand, graduating from high school is substantively important enough that even a suggestion that youth substitute work for school is worth additional attention. Appendix Section A.4 provides evidence that this substitution is real. There we use outcomes measuring work and school jointly to confirm that it is the same group of people driving the increase in work and decline in school progress (both on-time graduation and overall persistence). We also use GPA in the pre-randomization year to focus attention on students who are most likely to be at the margin of graduating. For below-median GPA students, the letters are associated with a significant decline in year 1 enrollment and GPA, as well as a substantively large and significant decline in on-time graduation (a decrease of 5.7 percentage points, or 8 percent). These students also show the largest employment change on the extensive margin.

We should use some caution in interpreting these results, since they involve an exploratory subgroup split among the already-reduced education sample. Nonetheless, all indications are that students who are close to the margin of failing to graduate take longer to complete high school, driven by those who shift from not working to working when they receive the letter of recommendation. Longer follow-up that includes 5th- and 6th-year graduation outcomes for the youngest cohorts is needed to know for sure if this group will eventually return to school.

5 Discussion

Sending youth a few copies of a recommendation letter and an email with a link to that letter improves their labor market trajectories. Short-term employment rates increase by 4.5 percent, while earnings effects continue to grow over 4 years, generating a cumulative 4.9 percent increase. These are big changes for a small intervention, providing new evidence that there are information frictions in the U.S. youth labor market that credible signals can overcome. The lack of differences in job-seeking behavior among treatment youth, other than using the letter itself, suggests that employers are the ones responding to the additional information contained in the letters. And given that higher performing youth get a larger labor market benefit from the letters, it seems that employers are successfully updating their beliefs about which applicants are likely to be productive workers.

We also find that recommendation letters may lead to a decline in on-time graduation, driven by substitution toward work among lower-achieving students. To assess the welfare implications of this substitution, we would need to make some strong assumptions about how long the increase in earnings will last and how that compares to the costs of additional time spent in high school. It seems likely that the net effect may not be beneficial, especially if, once the whole sample is old enough to observe final graduation outcomes, any subgroup has left school entirely. Reducing employment frictions is most likely to have a net benefit for those who have already finished their high school careers, or at least are not currently on the margin of failing to graduate. Current policy efforts to provide subsidized, year-round employment opportunities for high school students might therefore benefit from careful targeting.

For those not on the margin of on-time graduation, it is clear that providing credible signals about existing skills benefited the individual youth in our study. This is an important insight for social programs looking to help young people capitalize on their training or skills, and potentially for other disadvantaged groups facing employer uncertainty in the labor market (e.g., the formerly incarcerated). Finding additional ways to provide personalized information about an individual worker's strengths could help improve labor market outcomes among low-income, largely minority individuals like those in our study.

That said, it is crucial to emphasize that improvements among individuals are not the same as increases in net welfare. While the small size of our control group relative to the whole labor market makes within-study spillovers implausible, we do not have the data to assess the broader societal impact of providing some youth with letters. It is not clear whether the reduction in uncertainty decreased vacancies that would otherwise have been left open, as they appear to in an online marketplace (Pallais 2014), or whether other workers were displaced (and if so, whether the resulting hires were more or less productive than the potentially displaced workers).

The possibility of displacement, along with the local nature of our effects, means that it is not clear whether scaling up the provision of credible signals would generate welfare gains. Our effects are specific to the youth who receive letters when survey responses are voluntary and responses are positive enough. Effects may differ if a broader population of supervisors or young people were pushed into participating.[41]

41. It is difficult to say from the observable differences in youth across the opened and unopened surveys whether effects would be bigger or smaller if supervisors were forced to fill out the surveys. The unopened surveys contained more White youth, for whom we observe smaller labor market effects. But they also had more youth already out of high school, which could diminish graduation crowd-out, and more youth with work experience prior to SYEP, who have directionally larger point estimates on employment and earnings,

Of course, there are several conditions under which a scaled-up version could be beneficial, even with no net employment increase and full displacement. It is possible that widespread skill signals could generate efficiency gains by helping employers and employees find better matches. Alternatively, policymakers with preferences for equity might value transferring job opportunities to those farther down the income distribution or to historically marginalized groups. Finally, there could also be general equilibrium effects on the supply side; if young people understand that they may receive helpful recommendation letters, they may work harder in their jobs, generating additional productivity as well as better letters to which future employers will respond more positively.

These general equilibrium questions are an important avenue for future work. Further research into the precise way employers update their beliefs, substitute across workers, or change the number of employees in response to reduced uncertainty might be particularly productive next steps in assessing the most effective way to leverage our findings into welfare-enhancing labor market policies. But the key conclusion from our experiment—that information frictions reduce long-term earnings trajectories, even in high-information settings like the U.S.—does not rest on anticipating the effects of signals at scale in equilibrium. Documenting the existence of these frictions is enough to confirm the role of employer uncertainty in the labor market, as well as to raise key questions about the way in which this uncertainty might be limiting the success of active labor market programs and other efforts to improve the labor market prospects of young people.

see Appendix Section A.3.

References

Abebe, Girum, A Stefano Caria, Marcel Fafchamps, Paolo Falco, Simon Franklin, and Simon Quinn. 2021. "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City." *The Review of Economic Studies* 88 (3): 1279–1310.

Abel, Martin, Rulof Burger, and Patrizio Piraino. 2020. "The Value of Reference Letters: Experimental Evidence from South Africa." *American Economic Journal: Applied Economics* 12 (3): 40–71.

Agan, Amanda, and Sonja Starr. 2018. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *The Quarterly Journal of Economics* 133 (1): 191–235.

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

Altonji, Joseph G, and Charles R Pierret. 2001. "Employer learning and statistical discrimination." *The Quarterly Journal of Economics* 116 (1): 313–350.

Autor, David. 2001. "Why Do Temporary Help Firms Provide Free General Skills Training?" *The Quarterly Journal of Economics* 116 (4): 1409–1448.

Bandiera, Oriana, Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, and Anna Vitali. 2021. *The Search for Good Jobs: Evidence from a Six-year Field Experiment in Uganda.* SSRN Scholarly Paper ID 3910330. Rochester, NY: Social Science Research Network.

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

Bassi, Vittorio, and Aisha Nansamba. 2021. "Screening and Signalling Non-Cognitive Skills: Experimental Evidence from Uganda." *The Economic Journal* 132 (642): 471–511. ISSN: 0013-0133.

Baum, Charles L., and Christopher J. Ruhm. 2016. "The Changing Benefits of Early Work Experience." *Southern Economic Journal* 83 (2): 343–363.

Becker, Gary S. 1964. *Human capital: A theoretical and empirical analysis, with special reference to education.* University of Chicago Press.

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

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014a. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29–50. ISSN: 0895-3309.

———. 2014b. "Inference on Treatment Effects after Selection among High-Dimensional Controls†." *The Review of Economic Studies* 81 (2): 608–650. ISSN: 0034-6527.

Belot, Michele, Philipp Kircher, and Paul Muller. 2019. "Providing advice to jobseekers at low cost: An experimental study on online advice." *The Review of Economic Studies* 86 (4): 1411–1447.

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

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

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

Carranza, Eliana, Robert Garlick, Kate Orkin, and Neil Rankin. 2020. *Job Search and Hiring with Two-sided Limited Information about Workseekers' Skills.* W.E. Upjohn Institute.

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

Davis, Jonathan M.V., and Sara B. Heller. 2020. "Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs." *The Review of Economics and Statistics* 102 (4): 664–677.

DellaVigna, Stefano, Jörg Heining, Johannes F Schmieder, and Simon Trenkle. 2022. "Evidence on job search models from a survey of unemployed workers in germany." *The Quarterly Journal of Economics* 137 (2): 1181–1232.

Doleac, Jennifer L., and Benjamin Hansen. 2020. "The Unintended Consequences of "Ban the Box": Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden." *Journal of Labor Economics* 38 (2): 321–374.

Donovan, Kevin, Will Jianyu Lu, and Todd Schoellman. 2020. *Labor market dynamics and development.* Technical report. JSTOR.

Farber, Henry S, and Robert Gibbons. 1996. "Learning and wage dynamics." *The Quarterly Journal of Economics* 111 (4): 1007–1047.

Freeman, Richard Barry, and Harry J Holzer. 1986. *The black youth employment crisis.* University of Chicago Press.

Freeman, Richard Barry, and David A Wise. 1982. *The youth labor market problem: Its nature, causes, and consequences.* University of Chicago Press.

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 (1): 423–460.

Gonzalez, Francisco M, and Shouyong Shi. 2010. "An equilibrium theory of learning, search, and wages." *Econometrica* 78 (2): 509–537.

Heinrich, Carolyn J, and Harry J Holzer. 2011. "Improving education and employment for disadvantaged young men: Proven and promising strategies." *The Annals of the American Academy of Political and Social Science* 635 (1): 163–191.

Heller, Sara B, and Judd B Kessler. Forthcoming. "How to allocate slots: The market design of Summer Youth Employment Programs." *Fair by Design: Economic Design Approaches to Inequality, Eds. S.D. Kominers and A. Teytelboym.*

Hoffman, Mitchell, Lisa B. Kahn, and Danielle Li. 2018. "Discretion in Hiring." *The Quarterly Journal of Economics* 133 (2): 765–800.

Holzer, Harry J. 1988. "Search method use by unemployed youth." *Journal of Labor Economics* 6 (1): 1–20.

Hoynes, Hilary, Douglas L Miller, and Jessamyn Schaller. 2012. "Who suffers during recessions?" *Journal of Economic Perspectives* 26 (3): 27–48.

Jovanovic, Boyan. 1979. "Job matching and the theory of turnover." *Journal of Political Economy* 87 (5, Part 1): 972–990.

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, and Fabian Lange. 2014. "Employer learning, productivity, and the earnings distribution: Evidence from performance measures." *The Review of Economic Studies* 81 (4): 1575–1613.

Kahn, Lisa B. 2010. "The long-term labor market consequences of graduating from college in a bad economy." *Labour Economics* 17 (2): 303–316.

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 (1): 83–119.

Le Barbanchon, Thomas, Diego Ubfal, and Federico Araya. Forthcoming. "The Effects of Working while in School: Evidence from Uruguayan Lotteries." *American Economic Journal: Applied Economics.*

Lesner, Rune Vammen, Anna Piil Damm, Preben Bertelsen, and Mads Uffe Pedersen. 2022. "The Effect of School-Year Employment on Cognitive Skills, Risky Behavior, and Educational Achievement." *Economics of Education Review* 88:102241.

Modestino, Alicia Sasser. 2019. "How Do Summer Youth Employment Programs Improve Criminal Justice Outcomes, and for Whom?" *Journal of Policy Analysis and Management* 38 (3): 600–628.

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.

Mortensen, Dale T, and Christopher A Pissarides. 1999. "New developments in models of search in the labor market." *Handbook of Labor Economics* 3:2567–2627.

Neumark, David. 2002. "Youth Labor Markets in the United States: Shopping Around vs. Staying Put." *Review of Economics and Statistics* 84 (3): 462–482.

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 (1): 1–29.

Pallais, Amanda. 2014. "Inefficient Hiring in Entry-Level Labor Markets." *American Economic Review* 104 (11): 3565–3599.

Pallais, Amanda, and Emily Glassberg Sands. 2016. "Why the Referential Treatment? Evidence from Field Experiments on Referrals." *Journal of Political Economy* 124 (6): 1793–1828.

Papageorge, Nicholas W., Seth Gershenson, and Kyung Min Kang. 2020. "Teacher Expectations Matter." *The Review of Economics and Statistics* 102 (2): 234–251.

Rosenthal, Robert, and Lenore Jacobson. 1968. "Pygmalion in the classroom." *The Urban Review* 3 (1): 16–20.

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

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.

Stanton, Christoper T., and Catherine Thomas. 2016. "Landing the First Job: The Value of Intermediaries in Online Hiring." *The Review of Economic Studies* 83 (2): 810–854.

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

Vammen Lesner, Rune, Anna Piil Damm, Preben Bertelsen, and Mads Uffe Pedersen. 2022. "The Effect of School-Year Employment on Cognitive Skills, Risky Behavior, and Educational Achievement." *Economics of Education Review* 88. ISSN: 0272-7757.

Van der Berg, Servaas. 2007. "Apartheid's enduring legacy: Inequalities in education." *Journal of African economies* 16 (5): 849–880.

Tables and Figures

Table 1: Descriptive Statistics

	Control Mean	Treatment Mean	Test of Difference
N	21,695	21,714	
Age	17.2	17.2	0.641
Male	0.427	0.427	0.991
Black	0.409	0.411	0.805
Hispanic	0.289	0.289	0.944
Asian	0.129	0.130	0.734
White	0.124	0.125	0.756
Other Race	0.049	0.045	0.080
In High School	0.755	0.751	0.339
HS Graduate	0.044	0.042	0.202
In College	0.173	0.180	0.081
Not in UI Data	0.006	0.007	0.209
Never Employed Pre-SYEP	0.450	0.457	0.125
Ever Worked, Year -4	0.153	0.149	0.225
Earnings, Year -4	303	311	0.576
Ever Worked, Year -3	0.266	0.266	0.866
Earnings, Year -3	574	575	0.931
Ever Worked, Year -2	0.437	0.435	0.648
Earnings, Year -2	1052	1031	0.370
Ever Worked, Year -1	0.965	0.966	0.699
Earnings, Year -1	2334	2325	0.757
No Education Match	0.126	0.123	0.359
In HS Sample	0.454	0.454	0.938
Joint F-Test	$F(24, 41632) = 1.23$, $p=.201$		

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

Table 2: Labor Market Effects

Year	1	2	3	4	Cumulative
	Panel A: Employment				
ITT	0.0128***	0.0059	0.0029	0.0011	0.003
	(0.0041)	(0.0041)	(0.0043)	(0.0044)	(0.0025)
CM	0.701	0.720	0.650	0.682	0.922
Sent Letter (IV)	0.0316***	0.0145	0.0077	0.0028	0.0071
	(0.0102)	(0.0102)	(0.0107)	(0.0108)	(0.0062)
CCM	0.697	0.728	0.662	0.697	0.924
	Panel B: Earnings				
ITT	57.13	106.12	156.23	243.92*	531.42*
	(49.29)	(77.61)	(101.99)	(136.59)	(297.78)
CM	3594	6005	7457	10057	27141
Sent Letter (IV)	143.47	273.27	399.03	619.14*	1357.25*
	(121.71)	(191.76)	(252.05)	(337.61)	(735.64)
CCM	3764	6215	7579	10309	27943
	Panel C: Earnings, Winsorized at 99th Percentile				
ITT	57.96	104.37	128.83	214.72*	544.52**
	(43.16)	(71.94)	(96.65)	(128.38)	(277.26)
CM	3532	5925	7378	9927	26852
Sent Letter (IV)	149.02	267.11	330.4	546.06*	1348.83**
	(106.66)	(177.76)	(238.80)	(317.24)	(685.56)
CCM	3682	6132	7554	10239	27661

Notes: N = 43,409. Panel B shows raw earnings with a single outlier (>$3 million in earnings in one quarter) top-coded to next highest earnings in data. Winsorization in Panel C recodes each quarter's highest earnings to the 99th percentile of all quarterly earnings in a given year before summing to yearly totals. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are 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	Avg Spell Length (Spell 1-3)	Time to First Qtr Worked
ITT	0.062	0.019	0.010	0.047**	-0.103***
	(0.041)	(0.022)	(0.022)	(0.024)	(0.030)
CM	7.28	3.43	3.72	3.13	3.01
Sent Letter (IV)	0.152	0.046	0.025	0.117**	-0.254***
	(0.101)	(0.054)	(0.054)	(0.058)	(0.074)
CCM	7.53	3.43	3.71	3.24	3.01
N	43409	43409	40088	40088	40088

Notes: Spells are defined as consecutive quarters with earnings from same employer. The third through fifth columns condition on having at least one spell. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 4: Job Application Effects

	Clicked Link	Applied	Checked Selective Job Box	Uploaded Any File	Included Letter of Rec
ITT	-0.007	-0.006	-0.010	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)	-0.020	-0.019	-0.027	0.006	0.024***
	(0.024)	(0.022)	(0.017)	(0.018)	(0.007)
CCM	0.138	0.123	0.082	0.065	0.009

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

Table 5: Relationship between Ratings and Other Outcomes among Controls

Dependent Variable:	Cumulative Earnings		GPA Year 1		Graduated or Still Attending	
	Panel A: Full Sample Control Group					
Rating	1918.67***	971.70***				
	(171.17)	(153.14)				
Mean	27,243					
N	15487					
	Panel B: Education Sample Control Group					
Rating	792.77***	607.27***	1.80***	0.30***	0.027***	0.005**
	(170.94)	(164.50)	(0.10)	(0.06)	(0.003)	(0.002)
Mean	18,394		80.14		0.855	
N	7053		6532		7053	
Covariates	No	Yes	No	Yes	No	Yes

Notes: Coefficients from regressing each dependent variable on supervisor rating in the control group. Earnings and school persistence measured across 4 post-randomization years. Regressions include every control individual with a non-missing supervisor rating who is part of the the main sample (Panel A) or education sample (Panel B). Columns marked as having covariates include all the available baseline covariates for each sample listed in Appendix Section A.5. Standard errors clustered on individual are in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 6: Labor Market Effects for Youth with High and Low Supervisor Ratings

	First Stage	Y1	Y2	Y3	Y4	Cumulative
				Employment		
ITT, Low Ratings		0.0247*	-0.015	-0.0169	0.0096	0.0001
		(0.0133)	(0.0134)	(0.0139)	(0.0141)	(0.0080)
ITT, High Ratings		0.013	0.0237***	0.0157*	0.0067	0.0088*
		(0.0087)	(0.0087)	(0.0092)	(0.0093)	(0.0052)
P-value, test of diff.		0.463	0.015	0.051	0.865	0.361
CM, Low		0.673	0.721	0.657	0.669	0.924
CM, High		0.715	0.720	0.656	0.687	0.925
	First Stage					
IV, Low Ratings	0.3301***	0.0747*	-0.0454	-0.0511	0.0291	0.0002
	(0.0103)	(0.0404)	(0.0405)	(0.0422)	(0.0428)	(0.0242)
IV, High Ratings	0.8108***	0.0161	0.0292***	0.0194*	0.0083	0.0107*
	(0.0057)	(0.0108)	(0.0108)	(0.0114)	(0.0115)	(0.0064)
P-value, test of diff.	0.000	0.161	0.075	0.107	0.639	0.675
CCM, Low		0.613	0.757	0.688	0.668	0.917
CCM, High		0.713	0.717	0.657	0.687	0.924
				Earnings, Winsorized at 99th Percentile		
ITT, Low Ratings		25.60	-186.00	-191.12	-58.83	-355.69
		(121.61)	(207.51)	(266.56)	(347.72)	(758.53)
ITT, High Ratings		92.02	323.69**	222.78	562.34*	1239.98*
		(92.97)	(161.62)	(219.37)	(295.36)	(633.15)
P-value, test of diff.		0.664	0.053	0.231	0.173	0.106
CM, Low		3083	5369	6474	8361	23323
CM, High		3679	6205	7918	10880	28811
	First Stage					
IV, Low Ratings	0.3301***	75.62	-575.42	-586.89	-184.04	-1099.11
	(0.0103)	(368.07)	(628.68)	(807.87)	(1053.00)	(2298.22)
IV, High Ratings	0.8108***	112.86	394.95**	266.12	687.63*	1517.59*
	(0.0057)	(114.66)	(199.31)	(270.62)	(364.26)	(780.87)
P-value, test of diff.	0.000	0.923	0.141	0.317	0.434	0.281
CCM, Low		3115	5976	7047	8891	24887
CCM, High		3589	6045	7830	10785	28342

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

Table 7: Education Effects

	Ever Enrolled Y1	% Enrolled Days Present Y1	GPA Y1	Credits Attempted Y1-4	% Credits Earned Y1-4	Graduated On-Time	Ever Graduated	Graduated or Still Attending	On-time College
ITT	-0.002	0.001	-0.130	0.063	0.003	-0.007*	-0.001	-0.004	-0.004
	(0.003)	(0.003)	(0.099)	(0.100)	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)
CM	0.946	0.829	80.13	18.96	0.818	0.785	0.834	0.851	0.672
IV	-0.005	0.003	-0.302	0.194	0.006	-0.016*	-0.002	-0.009	-0.010
	(0.007)	(0.007)	(0.237)	(0.241)	(0.007)	(0.010)	(0.010)	(0.010)	(0.013)
CCM	0.957	0.849	81.75	18.49	0.844	0.828	0.865	0.883	0.717
N	19714	19714	18237	19714	19714	19714	19714	19714	17810

Notes: Analysis is conducted on all those expected to be observed based on pre-program grade of enrollment (see text for details). Percent credits earned is number earned divided by number attempted. Credits attempted and percent credits earned equal 0 for those not in school. 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 adds any 5th- and 6th-year graduation observed during the follow-up period. Graduated or still attending equals 1 if student either graduated or has positive days attended in most recent academic year. College enrollment is only measured within 6 months after a student's on-time graduation date, regardless of graduation status. CM shows control means; CCM shows control complier means. Regressions include baseline covariates. Standard errors clustered on individual are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Figure 1: Experimental Flow Chart

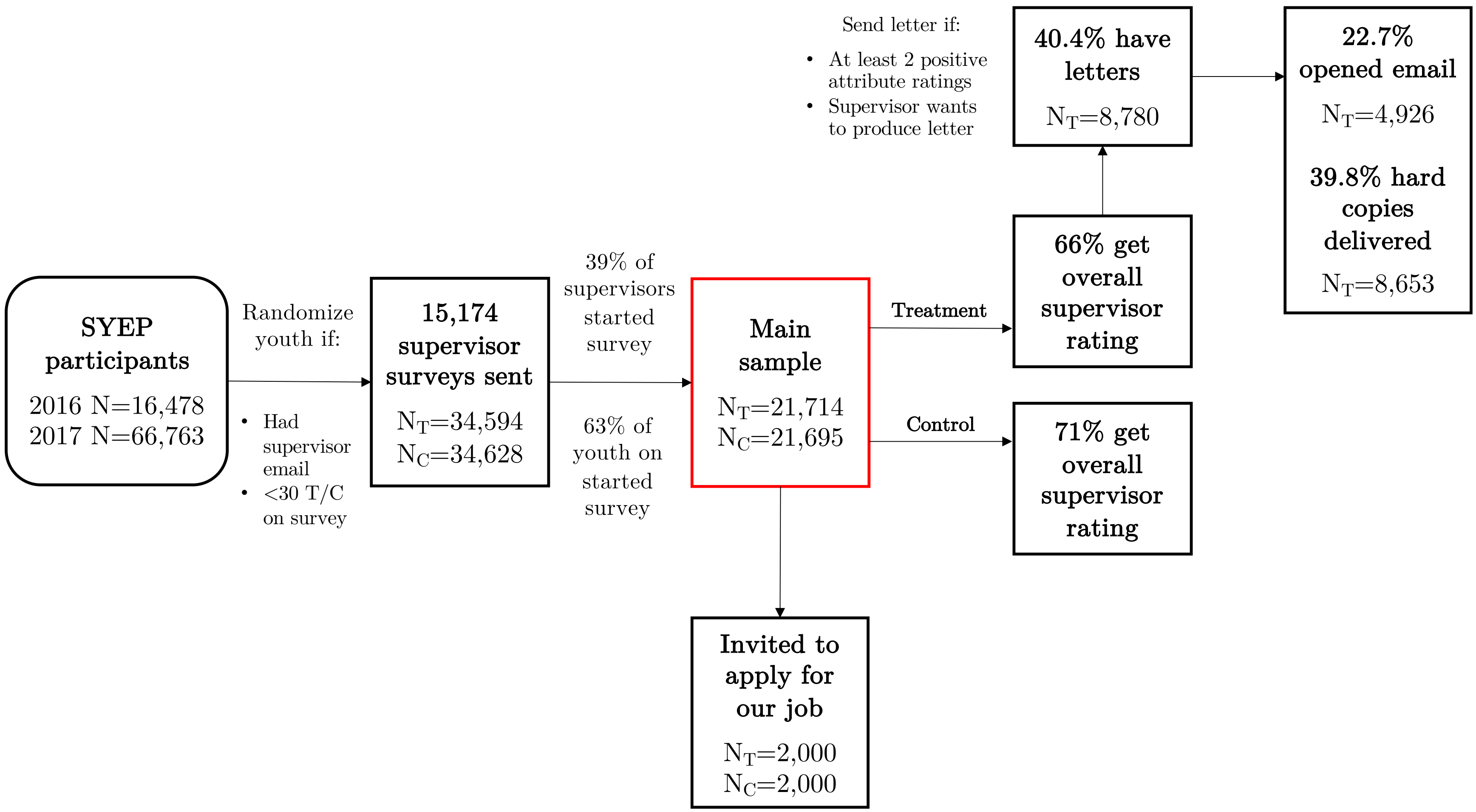
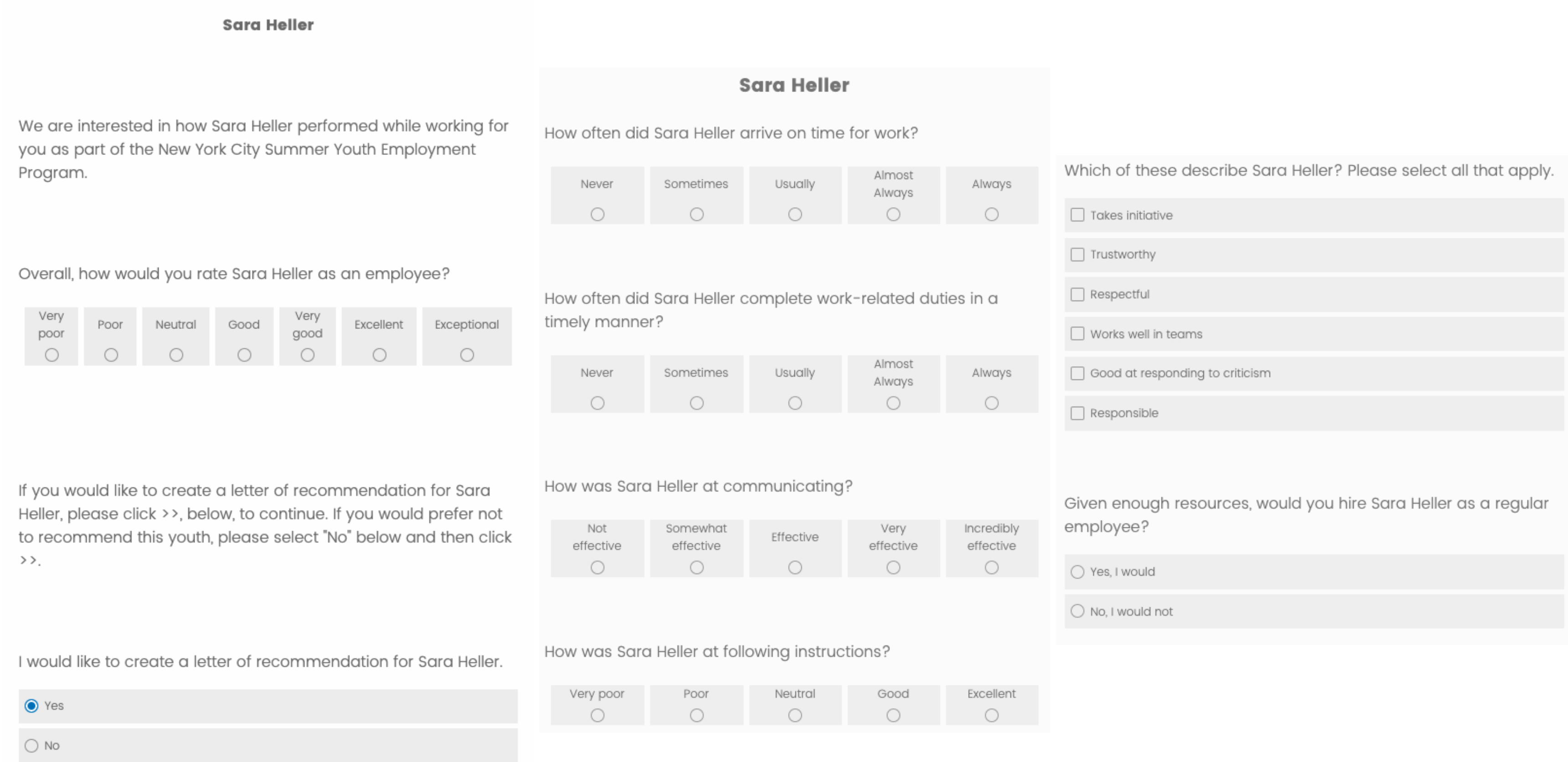


Figure 2: Screenshots about Treatment Youth on Supervisor Survey



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

Figure 3: Example Letter of Recommendation

NYC
Department of
Youth & Community
Development

November 1, 2017

To Whom It May Concern:

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

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

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

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

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

Sincerely,

Judd Kessler
The Wharton School

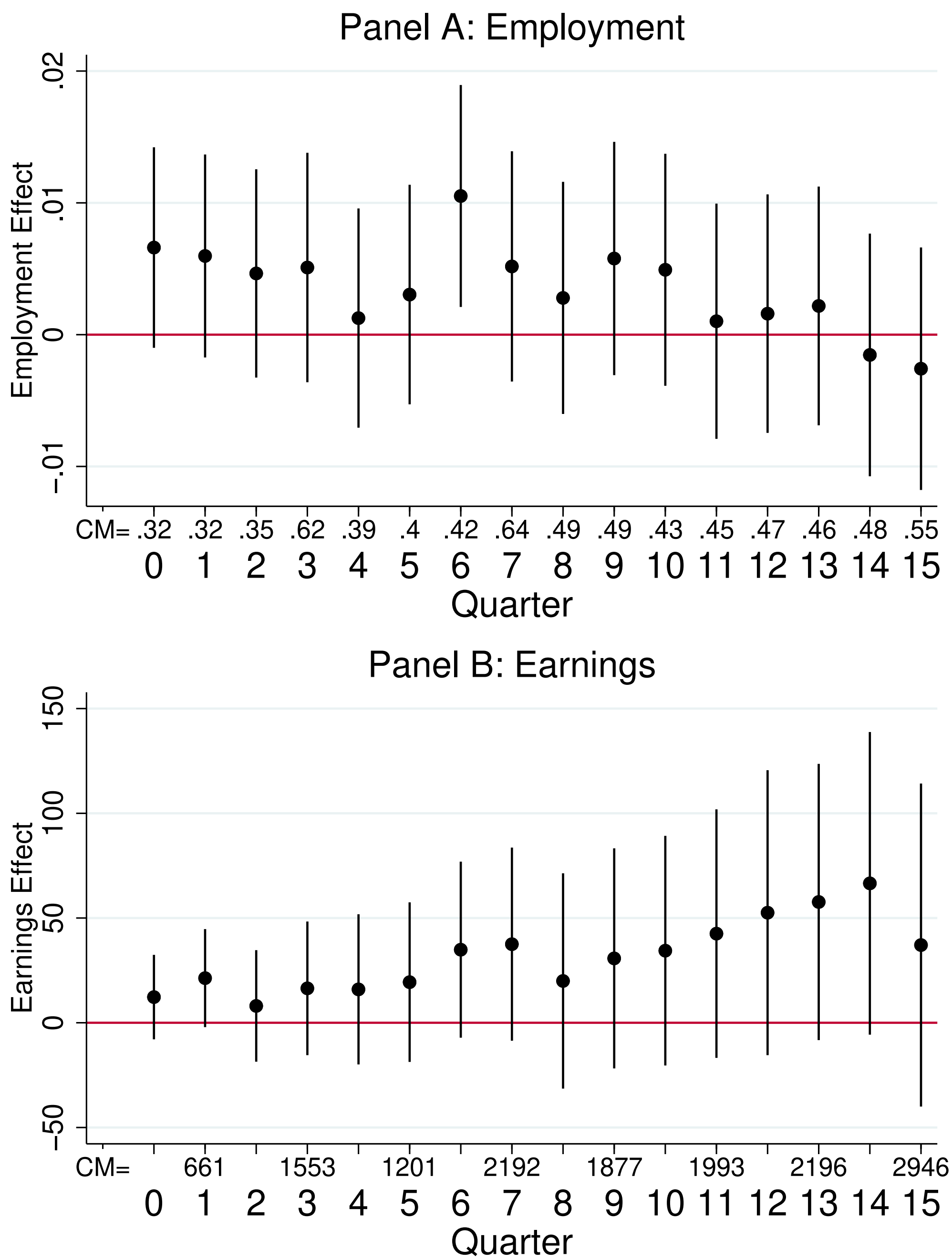
The New York City Department of Youth and Community Development (DYCD) invests in a network of community-based organizations and programs to alleviate the effects of poverty and to provide opportunities for New Yorkers and communities to flourish.

Empowering Individuals • Strengthening Families • Investing in Communities

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

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

Figure 4: Labor Market Effects by Quarter



Notes: Figure shows intent-to-treat effects on employment and earnings (winsorized at the 99th percentile) by quarter, with 95 percent confidence intervals calculated from standard errors clustered on individual. CM below axis displays the control mean in that quarter. All effects from regressions that include baseline covariates. Letters were distributed in Quarter 0. Quarters 0–3 comprise year 1, 4–7 are year 2, 8–11 are year 3, and 12-15 are year 4. Summers (July–September) are quarters 3, 7, 11, and 15.