Pre-Analysis Plan for "The Effects of Letters of Recommendation for Summer Youth Employment Program Participants"

Sara Heller, University of Michigan
Judd B. Kessler, University of Pennsylvania

1. Introduction

1.1 Background

In the wake of the Great Recession, youth employment has been recovering much more slowly than adult employment. Youth employment rates in the summer, when teenagers are most likely to be working, are still hovering near their 60-year low (Bureau of Labor Statistics 2018). Unemployment is disproportionately high among minority youth. In 2011, 40 percent of African-American young adults were jobless year-round compared with only 24 percent of whites (Sum et al. 2014). This level of unemployment among youth is likely to generate long-term harm; a range of evidence suggests that labor force attachment during adolescence and young adulthood affects employment and wages for decades (Kahn 2009; Neumark 2002; Oreopolous et al. 2012).

Policymakers have long been interested in using job training and placement programs to improve youth employment. However, there are few well-evaluated employment interventions that improve youths' labor market outcomes (Lalonde 2003). Those that succeed in doing so — programs like Job Corps, the National Guard ChalleNGe, and Year Up — involve lengthy and intensive interventions of a year or more, often with a residential component (Fein and Hamadyk 2018; Millenky et al. 2011; Roder and Elliott 2014; Schochet, Burghardt and McConnell 2008). The benefits of such intensive programming do not always exceed the costs.[1]

Yet recent evidence on one type of low-intensity intervention — summer youth employment programs (SYEPs) — suggests that it can have large effects on youth outcomes. Random-assignment studies from Chicago, New York City, and Boston suggest that 6-8 week summer job programs can have large and lasting impacts on violent offending (a 43 percent decline in violent crime arrests), incarceration (a 10 percent decline in incarceration in state prison), and mortality (a 20 percent decline in mortality), and estimates suggest that their net social benefits exceed their costs (Davis and Heller 2017a; Gelber, Isen and Kessler 2014; Heller 2014; Modestino 2017). Yet despite these successes on other outcomes, administrative records show no improvement in future employment outcomes from these programs on average, only tiny increases (on the order of 1-2 days) in school attendance, and no effect on college enrollment (Gelber, Isen, and Kessler 2014; Heller 2014; Leos-Urbel 2014). Indeed, in New York City, program participation is associated with slightly lower wages in the three years after the

[1] Job Corps, which usually involves an average residential stay of about 8 months, has costs higher than the benefits it generates (Schochet, Burghardt and McConnell 2008). The National Guard ChalleNGe, a pseudo-military education program with residential and non-residential components over more than a year, probably does generate benefits in excess of costs, though it depends how program effects develop over time (Perez-Arce et al., 2012). Year Up, which provides 6 months of classroom training followed by a 6-month internship, increases wages but not employment rates in the year after the program (Fein and Hamadyk 2018; Roder and Elliott 2014). Not enough time has passed to conduct a cost-benefit analysis.

program, seemingly due to participants' returning to low-earning jobs in the short-term following the program's conclusion (Gelber, Isen, and Kessler 2014).

Finding no positive impact to employment outcomes among SYEP participants is somewhat puzzling. Early work experience is associated with better future employment outcomes, potentially because it develops work and soft skills, a job history, and connections to employer networks (Mroz and Savage 2006; Ruhm 1995). It seems logical that providing summer jobs, especially to a population that does not often find work on their own, should do the same. It is possible that building connections to employer networks and teaching work and "soft" skills simply require more time than is available over the summer. It is also possible that whatever non-participants do in place of a SYEP is as or more constructive than the program itself. However, we investigate a third hypothesis: that information frictions constrain the labor market impacts of SYEPs. Such frictions could arise both on the "demand side" (i.e., employers) or the "supply side" (i.e., SYEP youth). On the demand side, future potential employers may not understand what youth learned in a SYEP, have negative beliefs about who chooses to participate in a SYEP, or have more general negative stereotypes or beliefs about the minority youth who tend to participate in these programs. Such employers might screen out SYEP participants before the interview stage. On the supply side, SYEP youth might not know that they have the skills necessary to hold certain jobs or might not feel confident enough about those skills to apply for jobs in certain industries or with certain employers. Either or both of the examples could prevent the skills developed in SYEPs from translating into positive labor market impacts.

1.2 Current Study

We aim to test whether providing additional information about New York City's SYEP participants in the form of personalized letters of recommendation can overcome these kinds of informational frictions and result in improved labor market or educational outcomes. Past research suggests that providing a small amount of information to adult decision-makers can be quite powerful: Moderately positive feedback in an online marketplace significantly improves an applicant's later employment outcomes (Pallais 2014), and employers appear to dramatically increase discrimination when they have less information about people (Agan and Starr 2016; Doleac and Hansen 2016). Studies in South Africa found that providing information about skills in a format that makes the information easy to share increased employment rates and earnings, and helped to close gender-based labor market gaps by providing better information on high-quality applicants (Abel 2017; Carranza et al. 2017).

Effects of information can also be seen in educational outcomes — even fictional positive information changes teachers' expectations in a way that improves student outcomes. In a classic study from the 1960s that has since been replicated many times, researchers found that elementary school students performed better on IQ tests if their teachers believed the students to be gifted, even if that belief had no basis in fact (Rosenthal and Jacobson 1968). Educators and policy makers have worked to harness this effect, known as the Pygmalion effect, to improve youth outcomes; we hypothesize that letters of recommendation — if shared with teachers and guidance counselors — could prove an effective means for accomplishing this. Additional letters of recommendation might also strengthen youths' college applications.

The potential of providing letters of recommendation to SYEP youth is underscored by evidence that SYEP employers are generally impressed by their SYEP employees. In Chicago's summer jobs program, for example, 76% of employers reported they would hire their youth immediately if a position were available, and over 80% of youth met or exceeded work-readiness standards (Bertrand and Heller *in progress*). By providing youth with a letter of recommendation about their strengths that they can use to shape how future potential employers and teachers view their summer jobs experience, we may be able to help youth capitalize on the strongly positive impressions that they made. If this turns out to be the case, our findings would have immediate policy implications: SYEPs across the country could easily and cheaply add letters of recommendation from SYEP employers as an enhancement to their programs.

We have implemented a randomized controlled trial to rigorously evaluate the impact of the letters of recommendation on labor market and education outcomes among NYC SYEP participants. In summers 2016 and 2017, we surveyed employers to collect personalized information, turned survey responses into letters of recommendation, provided the letters directly to youth, and encouraged youth to share them with potential employers as well as teachers and guidance counselors. In summer 2017, we also invited a subset of study youth to apply to a job posting. This invitation allows us to assess how much of any potential labor market change comes from differences in job-seeking behavior versus changes in employer perceptions. It will also give us a measure of how often youth actually use their letters.

Because we have been uncertain about which data sources we would be able to obtain, we did not pre-specify our outcomes prior to conducting the study itself. The only data we have in-hand at the moment is information on letter distribution and job applications to our own posting, which remains blinded to treatment status. Because we now have a sense of what data will be available, we are posting our pre-analysis plan (PAP). However, the data are not yet in hand. If needed, as we discover more about data availability and data quality, we will revise the PAP prior to unblinding the data.

The remainder of this pre-analysis plan expands on the study design and implementation (section 2) and discusses our data collection and analysis plan (section 3).

2. Experimental Design

2.1 Main Intervention

To implement our "Letter of Recommendation Program," we worked with the New York City Department of Youth and Community Development (DYCD), which administers the largest SYEP in the country. In both a pilot year (2016) and the full roll-out year (2017), we randomly selected a treatment group of SYEP participants to be eligible to receive letters of recommendation, and a control group of SYEP participants to be ineligible to receive letters. We surveyed SYEP supervisors to collect ratings across performance dimensions about each SYEP participant they supervised. We asked numerous questions about each treatment youth and one overall rating question about each control youth. Positive responses to questions were turned into sentences to construct a letter of recommendation for treatment youth. Each participant in the treatment group who received enough positive responses to the survey questions received a personalized letter of recommendation from their supervisor. The research team distributed letters of recommendation to youth by USPS and by email, which allowed us to track whose letters were returned (by mail) and who clicked the link to access their letter (in the email).

In summer 2016, we conducted a large-scale pilot of this process among a selected subset of SYEP participants. We imposed sample eligibility restrictions to ensure that both youth and employer contact information were available, and that no employer would be asked to rate more than 30 treatment youth in detail or provide an overall rating for more than 30 control youth. This resulted in a sample size of about 13,000 SYEP participants. We randomly assigned about half of the participants to the treatment group and the remainder to control. For each treatment youth with a positive enough survey response to generate a letter, we sent an e-mail with a link to a pdf version of the letter and we mailed 5 hard copies of the letter to their home address. Both sets of communication included an explanation of what the letter was and an encouragement to use the letter in job applications and to share it with teachers and guidance counselors.

In summer 2017, we used our experience in the pilot to implement a full-scale RCT among approximately 70,000 SYEP participants. As with the 2016 pilot, we limited the sample to the subset of participants who had both youth and employer contact information available, and we ensured that a single supervisor was not asked to evaluate more than 30 treatment or 30 control youth in the survey. Consequently, our 2017 sample was roughly 56,000 youth. We again e-mailed and mailed copies of the letter and instructions to treatment youth who received positive enough responses to generate a recommendation.

2.2 Job Application

Should the letter of recommendation intervention affect labor market outcomes, we would like to understand more about the mechanisms through which it does so. To distinguish demand-side and supply-side responses to the letters of recommendation, we distributed a job posting to a subset of 2017 participants (n = 5,000). To prevent the opportunity from interfering with other job-seeking behavior (our main outcome of interest), the posting made clear that the job was for a short-term set of tasks. The application asked some experience questions, gave youth an opportunity to upload supporting material, and asked about their desire to apply for a slightly higher paying job that was more selective. We hired everyone who completed enough of the application. The job asked youth to share information on their labor market experiences; we also chose a subset of those who provided more support and experience in their application to complete an extra task for a higher wage (brainstorm ways that SYEP could be improved to better train participants to succeed in the job market and in applying to college). Youth were paid for the two-hour job with a pre-loaded debit card sent to them by mail.

Applications to the job posting will provide a window into an otherwise unobserved first stage of letter of recommendation use; we will observe who actually applies for the job and chooses to submit a letter of recommendation. A treatment-control difference in application rates would provide an estimate of a “supply response” (i.e., the letters of recommendation induce increased job-seeking, either from increased self-efficacy and motivation or from improved beliefs about the probability of being hired). A null result would be of interest as well: If there is not a substantial supply response, it would suggest any employment changes are due to a “demand response” (i.e., employers changing their hiring behavior because of the letter).

We plan to use responses to our job posting as a rough estimate of the unobserved first stage. Since the posting was actively advertised directly to SYEP youth and the application provided an opportunity to upload documents and supporting material, we expect rates of submitting the application and letter of recommendation to be higher than rates in the regular job market. As such, we expect letter usage in these applications will be an upper bound on letter of recommendation use in the regular labor market. Since our pseudo-first stage is likely to be an upper bound, the implied local average treatment effect (LATE) we estimate from scaling the ITT with this first stage will be a lower-bound on the true LATE.

3. Analysis Plan

3.1 Sample Definition

To preserve the integrity of random assignment, every youth who was included on an employer survey invitation will be part of the main analysis sample.

However, because of expired supervisor e-mail addresses or supervisor disinterest, not every employer actually received and clicked the link inviting them to take the survey. Treatment youth who were on one of the unopened surveys had no chance of receiving a letter of recommendation. This kind of "non-compliance" will reduce our statistical power by reducing the number of treatment youth who actually received letters. To help maximize power, we tracked which employers clicked the link in their survey invitations. We plan to analyze the subset of youth whose employers clicked their initial links separately — regardless of whether those employers finished the full survey or rated any particular youth. There is little reason to think that limiting our analysis to this subsample should interfere with the integrity of random assignment, because employers had no way of knowing which of their supervisees would be included in the survey, nor which would be treatment or control, from the email invitation. As such, employer click-through should be orthogonal to youth treatment status. Of course, it is always possible in a finite sample that employer click-through behavior has some small correlation with the treatment status of the youth in their survey. If employer click-through is also correlated with youth outcomes in our finite sample, any finite sample correlation could potentially bias our estimate. To guard against this possibility, we will check baseline balance on this subsample separately. But we pre-specify this subsample as a key subsample of interest, since we expect it will help with statistical power. We note that the treatment effect in this subsample will be specific to the population of youth whose employers would receive and click the invitation if invited, which might differ from the treatment effect on the broader sample of all youth randomized. However, since it is unlikely that any policy would force employers to respond to a survey, the click-through group is likely to be a policy-relevant subsample.

3.2 Analytical Method

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

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

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. The elements of X_{it-1} will be chosen using a machine learning covariate selection method that seeks to keep residual variance small, while also avoiding data mining and generating correct inference. We will also show results without covariates as a robustness check. Because we will not observe the first stage at an individual level for everyone, our focus will be on this ITT estimate.

In addition to our main ITT specifications, we plan to conduct 2-3 additional sets of analyses.

1. *Job posting analysis*
   We will test for treatment-control differences in job application rates and the use of a recommendation letter in the application. This analysis will only include the subset of youth who received an invitation to apply to our job. The subset was a random subset (stratified on treatment group) of youth from 2017 who were in a survey opened by a supervisor (n = 2,000 each for treatment and control) and youth who were not in a survey opened by supervisors (n = 500 each for treatment and control). The specification will be similar to our main ITT analysis.

   The results will generate an implied "first stage" — the treatment-control difference in use of a letter in a job application. In addition to helping us unpack the mechanisms behind any potential changes in labor market outcomes, we will also use the implied first stage to back out an implied LATE for those whose job-seeking behavior was changed by random assignment.

2. *Heterogeneity analysis*
   Although we do not anticipate having enough statistical power to be able to differentiate heterogeneous treatment effects, we will conduct some exploratory analysis based on the characteristics most likely to affect whether youth are actively looking for a job and the nature of the labor markets they face: age, school enrollment (if observable), race, gender, and neighborhood. Given the richness of the data, we anticipate leveraging machine learning approaches to heterogeneity as part of our exploration (e.g., Davis and Heller 2017b). However, due to confidentiality concerns of our data providers, we are unsure that we will be able to combine education data with our other outcomes, so the scope for principled data exploration may be limited. We also anticipate testing for heterogeneity across employer ratings.

3. *SYEP evaluation*
   Although our main focus is on the sample of youth who were included in one of our employer surveys (a subset SYEP participants), we are hoping to have enough data on all SYEP applicants to conduct an updated evaluation of NYC's SYEP as well. Recent impact evaluations stopped with the 2010 cohort (Valentine et al. 2017), so it is desirable to test whether program impacts are similar in more recent cohorts, which participate in a somewhat updated program and face a different labor market. It is not yet clear whether we will have all the data needed to do so. But if possible, we will run ITT and TOT analyses on the full sample of applicants. Specifications will be similar to those in

Gelber, Kessler and Isen (2017), in particular including the set of provider fixed effects needed to ensure that treatment is conditionally random.

3.3 Data

Note that since we do not yet have data sharing agreements in place, our posted outcomes are tentative. We will update the PAP once we finalize data-sharing agreements and better understand any data quality issues. All decisions will be made while blinded to treatment status.

We are applying for access to administrative data from the New York State Department of Labor (NYSDOL) and the New York City Department of Education (DOE) to evaluate the impact of the Letter of Recommendation Program on labor market and education outcomes. The labor market outcomes we plan to use are from NYSDOL's Unemployment Insurance (UI) records, which include quarterly reports of gross earnings and associated employer NAICS codes. The education records would include de-identified (scrambled identifiers unlinked to other study data) student records on enrollment, grades, attendance, discipline, graduation, and college enrollment. Some of our data providers may require us to collect some of these measures averaged across small groups to avoid observability issues that may arise with individual-level data. If so, this will require us to adjust our regression specification to address the averaged nature of the data. An alternative approach to preserve confidentiality may require us to keep labor and education data entirely separate, with identifiers scrambled to prevent re-merging (which would affect which covariates we could include in each set of outcome regressions). We intend to analyze labor market and education records for the three-year period following random assignment in one-year increments. We are also interested in criminal justice outcomes but are not confident in our ability to obtain the necessary data.

3.4 Primary Outcome

Our primary outcome of interest for evaluating the Letter of Recommendation intervention will be annual earnings in the UI data. If there are extreme outliers, we anticipate winsorizing the data prior to looking at the treatment status of the outliers. We will show the robustness of our results to different ways of handling skewness.

3.5 Secondary Outcomes

Employment:

- Binary indicator for any employment
- *For subset of sample invited to apply for our job*: indicator for having applied, indicator for having uploaded a letter of recommendation

Education:

- To reduce the number of tests, we will combine several measures of academic engagement and performance into a single index (see Kling, Leibman and Katz 2007; Anderson 2008) for youth who had not graduated prior to random assignment. This will include two measures of attendance: days attended and school persistence (an indicator for either having graduated or still being enrolled in school) and measures of performance: GPA and standardized test scores when available (NYS Regents). The index will be calculated by standardizing each measure and averaging these z-scores to get an overall index value. We will also show the outcomes separately, as well as in their original units, to better understand what is driving any changes in the index, adjusting for multiple testing. This may be particularly important given the possibility for effects on index elements that go in opposite directions (e.g., pulling marginal youth into the labor market might pull them out of school, but also increase GPAs among the youth remaining in school).

- Indicator for any post-secondary enrollment for the subpopulation who are at least 18 by the follow-up period

3.6 Exploratory Data Analysis

We will perform exploratory analyses to probe potential mechanisms that contribute to effects we find. We pre-specify some of these tests below. However, we expect other hypotheses may arise from our data analysis, and we may also obtain additional data sources. We will indicate in the paper the exploratory nature of these tests. In the case that we secure additional outcome measures, we will revise our PAP prior to unblinding the data.

- Number of jobs in a year and tenure of work at an employer, to measure job stability
- Distribution of industry type. Prior work has suggested that youth who participate in SYEP on average have lower short-term earnings in the years immediately following SYEP participation because they have an increased likelihood of being employed in industries that are typically associated with SYEP jobs (Gelber, Isen and Kessler 2017). Relative to the employment industries of all youth in the SYEP age range, SYEP placements tend to skew toward lower paying industries. It is possible that letters of recommendation from SYEP employers will help youth communicate their SYEP experience to potential employers in industries outside of those similar to their SYEP placement. To investigate this, we will compare the distribution of industries in which treatment and control youth work, based on the categories defined in Gelber, Isen and Kessler.
- Indicators for 4-year and 2-year college enrollment, to differentiate any changes in type of post-secondary education.
- Also see section 3.1 for planned heterogeneity analysis.

References

Abel, Martin, Rulof Buger, and Patrizio Piraino. 2018. "The Value of Reference Letters - Experimental Evidence from South Africa." Working Paper.

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. https://doi.org/10.1093/qje/qjx028.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–95. https://doi.org/10.1198/016214508000000841.

Bureau of Labor Statistics, "Employment and Unemployment Among Youth - Summer 2018," Technical Report August 2018. https://www.bls.gov/news.release/youth.htm.

Davis, Jonathan M.V., and Sara Heller. 2017a. "Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs." w23443. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w23443.

Davis, Jonathan M.V., and Sara B. Heller. 2017b. "Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs." *American Economic Review* 107 (5): 546–50. https://doi.org/10.1257/aer.p20171000.

Doleac, Jennifer, and Benjamin Hansen. 2016. "Does 'Ban the Box' Help or Hurt Low-Skilled Workers? Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden." w22469. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w22469.

Fein, David and Jill Hamadyk, "Bridging the Opportunity Divide for Low-Income Youth: Implementation and Early Impacts of the Year Up Program," OPRE Report #2018-65, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2018.

Gelber, Alexander, Adam Isen, and Judd Kessler. 2014. "The Effects of Youth Employment: Evidence from New York City Summer Youth Employment Program Lotteries." w20810. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w20810.

Heller, Sara B. 2014. "Summer Jobs Reduce Violence among Disadvantaged Youth." *Science* 346 (6214): 1219–23. https://doi.org/10.1126/science.1257809.

Kahn, Lisa B. 2010. "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy." *Labour Economics* 17 (2): 303–16. https://doi.org/10.1016/j.labeco.2009.09.002.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. “Experimental Analysis of Neighborhood Effects.” *Econometrica* 75 (1): 83–119. https://doi.org/10.1111/j.1468-0262.2007.00733.x.

LaLonde, Robert J., "Employment and training programs," in *Means-tested transfer programs in the United States*, Robert A. Moffitt, ed. (Chicago, IL: University of Chicago Press, 2003).

Millenky, Megan, Dan Bloom, Sara Muller-Ravett, and Joseph Broadus. 2012. “Staying on Course: Three-Year Results of the National Guard Youth Challenge Evaluation.” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2019770.

Modestino, Alicia Sasser. 2017. “Reducing Inequality Summer by Summer: An Analysis of the Short-Term and Long-Term Effects of Boston’s Summer Youth Employment Program.” Program evaluation. Boston, MA: The Mayor’s Office of Workforce Development and Northeastern University’s Dukakis Center for Urban and Regional Policy. https://s20289.pcdn.co/wp-content/uploads/2018/04/Reducing-Inequality-Summer-by-Summer_FINAL.pdf.

Mroz, Thomas A., and Timothy H. Savage. 2006. “The Long-Term Effects of Youth Unemployment.” *Journal of Human Resources* XLI (2): 259–93. https://doi.org/10.3368/jhr.XLI.2.259.

Neumark, David. 2002. “Youth Labor Markets in the United States: Shopping Around vs. Staying Put.” *Review of Economics and Statistics* 84 (3): 462–82. https://doi.org/10.1162/003465302320259475.

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. https://doi.org/10.1257/app.4.1.1.

Pallais, Amanda. 2014. “Inefficient Hiring in Entry-Level Labor Markets.” *American Economic Review* 104 (11): 3565–99. https://doi.org/10.1257/aer.104.11.3565.

Perez-Arce, Francisco, Louay Constant, David S Loughran, and Lynn A Karoly, "A cost-benefit analysis of the national guard youth ChalleNGe program," RAND Research Monograph, TR1193, (2012).

Roder, Anne, and Mark Elliot. 2014. “Sustained Gains: Year Up’s Continued Impacts on Young Adults’ Careers.” New York: Economic Mobility Corporation.

Rosenthal, Robert, and Lenore Jacobson. 1968. “Pygmalion in the Classroom.” *The Urban Review* 3 (1): 16–20. https://doi.org/10.1007/BF02322211.

Ruhm, Christopher J. 1995. “The Extent and Consequences of High School Employment.” *Journal of Labor Research* 16 (3): 293–303. https://doi.org/10.1007/BF02685757.

Schochet, Peter Z, John Burghardt, and Sheena McConnell. 2008. "Does Job Corps Work? Impact Findings from the National Job Corps Study." *American Economic Review* 98 (5): 1864–86. https://doi.org/10.1257/aer.98.5.1864.