Essays in Education and Labor Economics

By

Jennifer A. Boden

M.A., University of Kansas, 2012

B.A., University of Northern Iowa, 2010

Submitted to the graduate degree program in Economics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____

Chair: Donna K. Ginther

____

Bozenna Pasik-Duncan

____

Dietrich Earnhart

____

David Slusky

____

Tsvetan Tsvetanov

Date Defended: 25 January 2019

The dissertation committee for Jennifer A. Boden certifies that this is the approved version of the following dissertation:

Essays in Education and Labor Economics

Co-Chair: Donna Ginther

Date Approved: 29 January 2019

Abstract

This paper explores treatment effects in three different contexts: an extended school year pilot program in an urban school district, a library summer reading program serving twenty-one urban school districts, and states passing the Equal Rights Amendment during the 1970s and 1980s.

Estimates generated using two years of intervention data along with difference-in-differences and nearest neighbor approaches show that a new extended school year (ESY) program in the North Kansas City School District (NKCS) improved mathematics MAP test scores in the 4th grade treatment group and mathematics as well as communication arts MAP test scores in the 3rd grade treatment group when compared to the corresponding control group by approximately .3 standard deviations. This is a large and statistically significant effect and indicates that the ESY intervention is improving achievement for these groups.

Summer learning loss has been well documented in the education literature. One attempt at combating this learning loss has been through the contributions of public libraries and their associated summer programs. Assessing program impact is difficult because public libraries cross the boundaries of multiple school districts. This study highlights a case study which outlines a data-driven, research method that can be used to measure program impact by linking incongruent sources of academic data. The paper discusses the challenges associated with working with data provided by multiple school districts and finds that participation in this particular summer reading program is associated with better outcomes for elementary school students.

The United States Constitution does not guarantee equal rights based on gender. The Equal Rights Amendment (ERA), passed by Congress in 1972, was intended to remedy this situation, but only thirty-five of the requisite thirty-eight states passed the law before its 1982 expiration. Making this topic relevant for today, the ratification of an unrelated amendment 203 years after its Congressional approval has given ERA supporters renewed hope of its eventual passage. In fact, in 2018 the ERA was passed by both Nevada and Illinois taking the total number of passing states to thirty-seven, just one state short of the number needed for ratification. The state by state variation in terms of ERA ratification affords the opportunity to examine the labor market effects of this legislation. Because the existence of the gender wage gap has been well-established in the literature, the findings may be helpful in constructing policy interventions that can eliminate it. In this study, we find that, relative to men and other women, white women in ERA states were more likely to enter the labor force, but non-white women were the beneficiaries of increased pay and hours of work.

Acknowledgements

I would like to express my deepest gratitude to my advisor, Donna Ginther. She has spent every day of this process modelling the type of teacher, mentor, and researcher to which I can aspire. I know in my heart that none of this would have been possible without her. She was a consistent and supportive presence at a time when I faced some personal challenges, and I am truly fortunate to have had my time at the University of Kansas intersect with hers.

In addition to my advisor, I would like to thank the other members of my committee: Bozenna Pasik-Duncan, Dietrich Earnhart, David Slusky, and Tsvetan Tsvetanov. They spent many hours serving on the committee itself, giving feedback, asking questions, and providing encouragement.

I am also grateful to Pat Oslund and Karin Chang. Pat has the spirit of a teacher and the patience of a saint. I am so thankful for her willingness to share her vast knowledge with me. I could not have finished this work without her guidance. Karin was so gracious in providing flexibility in the workplace and allowing me to work on projects that overlapped with my dissertation topics. She always had a smile and a positive outlook, and she opened the door for new areas of research for me.

I would like to extend my sincere appreciation to two of my dearest friends, Danyce Jacobs and Craig Montgomery. Danyce was my cheerleader when I first decided I wanted to pursue a PhD, and she has been with me every step of the way. She has listened to me explain my research, practice my conference presentations, and has guided me in learning to be a better teacher. Craig helped me carve out the time I needed to finish my work and had the wisdom to know when I needed a break. He was an endless stream of pep talks and hugs when my

frustrations grew, and he made sure that I always remembered why I chose this path. I am truly blessed to know each of them.

Finally, I would like to thank my children who do not even remember a time when their mother was not in school. They embraced a move to another state so that I could go to school, and they helped me the best that they could.

Chapter 1: An Evaluation of an Extended Elementary School Year Program in an Urban Setting was coauthored with Donna Ginther, Ph.D. and Patricia Oslund, M.A. *Chapter 2: Linking Incongruent Data Sources: A Case Study of a Summer Library Program* was coauthored with Meghan Ecker-Lyster, Ph.D. and Karin Chang, Ph.D. *Chapter 3: The Equal Rights Amendment and Labor Market Outcomes for Women* was coauthored with Donna Ginther, Ph.D.

Table of Contents

Chapter 1: An Evaluation of an Extended Elementary School Year Program in an Urban Setting

Introduction

The educational achievement gap perpetuates economic inequality. Compared with children from advantaged backgrounds, children from low-income families who receive free and reduced lunch, children of color, and English language learners are less likely to score as proficient in school assessments (Jencks and Phillips, 1998; Reardon, 2011), graduate from high school (Losen, 2004; Rumberger, 2011), and attend college (Bennett and Xie, 2003; Goldrick-Rab and Pfeffer, 2009; Bailey and Dynarski, 2011). These differences were identified in the 1960s (Coleman et al., 1966) but have remained largely intractable (Gamoran and Long, 2007).

Recognizing inequality in their students' outcomes, the North Kansas City Public School District (NKCS) extended the school year for two elementary schools in the district with the goal of narrowing the achievement gap. Beginning in the 2015-16 academic year, NKCS, a large urban district in the Kansas City, Missouri metropolitan area, added 31 days to the school calendar in two of the 21 elementary schools in the district. This Extended School Year (ESY) policy was enacted to reduce the underperformance of students in schools with high rates of poverty and high percentages of English language learners (ELL). These racially and ethnically diverse schools served 820 students and had over 75% of students enrolled in the free or reduced lunch program (FRL) in 2015, with 22% ELL students. The extended calendar will not be offset by shorter school days or more time on the same curriculum. To date, fewer than 150 elementary schools, primarily charter schools, have adopted an extended calendar (National Center on Time & Learning, 2014). The NKCS ESY policy provides a rare opportunity to learn how added instructional time will affect children and may potentially narrow the achievement gap. This study examines whether the ESY policy implemented by NKCS improved achievement of students in the treated schools relative to a control group of schools with students from similar SES backgrounds.

Despite the interest in alternative school year and school day schedules, the research base for such innovations is thin. Theoretically, additional instructional time via ESY or Extended School Day (ESD) schedules should improve students' achievement outcomes. The literature on summer learning loss and its solutions, on year-round schooling, and on extended learning time all address some aspects of our research, but the literature leaves important questions unanswered.

Summer learning loss has been well documented in the literature, and the consensus is this loss is exacerbated for low-income and other disadvantaged children (Cooper et al., 1996; Alexander, Entwisle, and Olson, 2001). In the same vein, literature on summer interventions shows well-structured programs benefit low-income participants (Kim and Quinn, 2013). This suggests that the enrichment aspect of the NKCS ESY will benefit students in the treated schools.

Year-round schooling is generally implemented by redistributing the existing number of instructional days (approximately 180) across the calendar, with short breaks several times per year. The literature suggests the alternative calendar alone does little to advance overall student achievement (McMillan, 2001; Crow and Johnson, 2010; Wu and Stone, 2010) and in some cases may actually result in lower test scores (Graves, 2011). Simply rearranging a fixed number of instructional days appears to have little academic benefit. We have identified no rigorous research on the impact of extended calendars in public schools. Thus, research questions remain about the impact of ESY policies, particularly related to low-income students in the public-school context.

A review of the literature on extended learning time examines 27 ESD programs and 28 ESY programs (Redd et al, 2012). Both types of programs increase the amount of classroom contact time. The authors point out that although there is there is suggestive evidence of positive impacts, methodological issues, particularly the lack of rigorous experimental or quasi-experimental designs, prevent the evidence from being conclusive. A second review (Patall, Cooper, and Allen, 2010) focuses on 15 *quantitative* studies of expanded time schools (expanded day, year, or both). This review also concludes that the research design of the generally is weak, and the literature is limited to examining test score outcomes. In contrast, a study by Angrist et al. (2012) featured an experimental research design. They studied the outcomes of students chosen by lottery for charter schools with an ESY. In addition to ESY, the charter school curriculum emphasized math and reading skills, selective teacher training, and strict behavioral standards. The authors find positive achievement results in reading and math, with gains concentrated on special needs students for those who won the lottery of attending the charter school compared to non-winners in traditional schools. It is unclear, however, whether gains are due to ESY or to other innovations in the charter school setting.

One way that some students are given the opportunity for extended instructional time in the U.S. is through the 21st Century Community Learning Centers. James-Burdumy, et al (2005) randomly assigned elementary school students to these centers and found that compared to other students, the attendees were not more likely to show high academic achievement.

While there is some evidence that extended learning time improves student outcomes, understanding the mechanism is an important goal of this research. How does extended time affect what goes on in the classroom? Case studies compiled by the National Center on Time and Learning provide insights into this important component of extended learning time. Kaplan and Chan (2011) find that extended school time has been used by schools to: a) optimize time for student learning and individualized instruction; b) reinforce core values and emphasize school and career readiness; and c) devote time to teacher improvement. Similarly, Kaplan, Chan, Farbman, and Novoryta (2014) examine the effect of expanded time on teacher collaboration and other practices in 17 high-performing extended time schools. They find that teachers spend significantly more time on professional development in extended time schools compared to traditional schools (40 percent compared with 20 percent). Teachers use expanded time for activities including collaborative lesson planning, embedded professional development, summer training, data analysis, and individualized coaching.

The North Kansas City School District Extended School Year

NKCS selected the two schools for ESY separated by several miles so alternative neighborhood schools could be available for families who chose to opt out of the extended-year option. This policy prescription will aid in our research design since we will be able to match ESY schools to geographically contiguous schools within the NKCS boundaries; these comparison schools have similar demographics to the ESY schools and serve about 1,100 students annually. Teachers were also allowed to opt in or out of these schools, and increased instructional time means higher salaries for teachers in ESY schools. For beginning teachers, 31 additional instructional days will increase salaries by over $6,000 per year, and for experienced teachers by over $12,000. NKCS expects these policy shifts will be revenue neutral because NKCS will increase the number of student attendance days, leveraging state funding from regular and summer programming to sustain this endeavor.

Data

All data for this analysis were provided by NKCS and the Missouri Department of Elementary and Secondary Education (DESE). *Student performance*, measured with individual-level state assessment scores for reading and math, were used to determine whether student achievement improved in ESY schools. Demographic characteristics were extracted from school district databases in order to determine whether ESY programs close the achievement gap between advantaged and disadvantaged students.

Our treatment group will be all students who attended the NKCS ESY elementary schools during both post-intervention academic years, 2015-16 and 2016-17, and attended any NKCS elementary school for academic year 2014-15. Our control group will consist of students enrolled in contiguous elementary schools within the NKCS district during the same time period. A comprehensive list of NKCS elementary schools and the corresponding treatment designations may be seen in *Appendix A*.

FRL and ELL percentages act as a proxy for percent disadvantaged students in a given school. FRL and ELL percentages for treatment, control, and other district elementary schools may be seen in *Figure 1*.

Starting in third grade, Missouri students are tested annually using the Missouri Assessment Program (MAP). Through eighth grade, these students are tested on their content knowledge in communication arts and mathematics[1]. Prior to testing, the State Board of Education, with input from various stakeholders, approves a set of achievement cut scores. These cut scores, which vary by grade and content area, serve as thresholds dividing student MAP scores into four levels of achievement: Below Basic, Basic, Proficient, and Advanced. Because MAP scores for individuals are expected to grow over time and can change scale from

[1] Fifth and eighth grade students are also tested on content knowledge in science.

one year to the next, practitioners often use these achievement levels as a consistent way to interpret student data. The four achievement levels along with a brief description of each may be seen in *Table 1*.

To get an aggregate view of proficiency, the four levels can be used to calculate proficiency rates for districts, schools, or even demographic groups. To do this, the four achievement levels are condensed into two. Students designated as Level 1 or 2 are deemed *Not Proficient*, and students designated as Level 3 or 4 are deemed *Proficient*.

$$Proficiency\ Rate = \frac{Number\ of\ students\ designated\ as\ Proficient}{Number\ of\ students\ designated\ as\ Proficient\ or\ Not\ Proficient}$$

Figures 2 and *3* illustrate the changes in proficiency rates for third through fifth graders in NKSD elementary schools in communication arts and mathematics, respectively, from 2009 to 2017. Elementary schools are clustered into three groups: treatment schools, control schools, and other schools. The two schools that experienced the ESY intervention starting in academic year 2015-16 make up the treatment group. The control schools are the three schools that were considered for the ESY intervention, due to their similarity to the treatment schools and geographic proximity, but ultimately did not receive it. The other schools group is made up of the remaining NKCSD elementary schools. The dotted lines denote changes in proficiency rates during the ESY intervention.

The figures have several notable characteristics that informed our methodology. Each of these characteristics will be discussed in turn.

Our approach for understanding the impact of the ESY intervention is analogous to what researchers use to understand medical treatments. In a medical treatment study, individuals are divided into a treatment and control group based on whether they exhibit the same health problem. For example, if researchers were testing the impact of a diabetes treatment, both the treatment and control group would be required to have diabetes. The treatment group allows us to observe the outcome of interest for individuals who are exposed to the intervention. The purpose of the control group is to serve as an approximation for the outcome of interest for the treated individuals if they were left untreated. The impact of the treatment on diabetes is found by comparing the average outcomes for those in the treatment group to those in the control group. If the treatment group has better health outcomes than the control group (e.g., lower blood sugar) then the treatment is considered effective.

In the case of the NKCSD, the treatment and control schools should have similar "pre-treatment" conditions, in this case low achievement as observed by the relatively low percentages of proficient students seen in *Figures* 2 and *3*. The fact that the treatment and control schools behave similarly over time regarding proficiency rates provides some evidence that we have identified an appropriate control group. The fact that both of these groups fall substantially below the other district schools means that the possibility of affecting change through the intervention and being able to detect that change is possible. In other words, it would be difficult to detect change as a result of an intervention implemented if the schools were already performing well.

According to the 2015 MAP technical report[2], the 2014 and 2015 MAP assessments are not comparable. Evidence of this can be seen in *Figures 2* and *3*. Changes made to the 2015 assessment are associated with a substantial increase in proficiency rates. Although the assessment change may be enough to explain the large increase in proficiency rates, the scale,

[2] https://dese.mo.gov/sites/default/files/asmt-gl-2015-tech-report.pdf accessed on May 29, 2018.

and thus the associated cut scores, changed as well. For example, a third-grade student scoring 628 or higher on the math assessment would be considered proficient in 2014, but a similar student would require a score of 2436 in 2015. Whether the 2014 and 2015 cut scores have the same underlying meaning is uncertain. Generally speaking, it is difficult to ascertain whether an outcome has changed as a result of an intervention when there is a major change in how that outcome is measured. As a result of the 2015 cut score and assessment changes, pre-intervention data are omitted for years prior to 2015.

Because the MAP scale scores change over time, scores were standardized into z-scores using statewide means and standard deviations for each grade and content area. This means that the outcome of interest is interpreted as the number of standard deviations a score falls from the corresponding statewide average. For example, a z-score of 0 indicates that there was no difference between the test score and the statewide average for the corresponding grade and content area. Positive z-scores indicate that students scored above the statewide average, and negative z-scores indicate that students scored below statewide averages.

Since students at different grade levels experienced different dosages of the intervention, we divided students into grade-related groups which we refer to as cohorts. A description of these groups may be seen in *Table 2*.

For illustrative purposes, the average standardized MAP Score for the first year of intervention are presented in *Table 3*. The only consistent pattern we see is that the other district schools outperform both the treatment and control groups in communication arts as well as mathematics. Cohort 1 in the treatment group had higher average Communication Arts scores than the control group, whereas the reverse was true for Cohorts 2 and 3. The treatment group in

Cohorts 2 and 3 had much larger mathematics scores than the control group. However, this was not the case for Cohort 1.

Table 3 can help give context to the z-scores. For example, if treatment school z-scores increase by .1, it means that the intervention has helped close the gap between treatment and other district schools. In fact, an increase of .1 for cohort 1 in communication arts, means that the intervention has more than doubled the average score for those schools (.086 + .1 = .186) and narrows the gap between treatment and other districts schools (original gap: .308-.086 = .222; new gap: .308-.186 = .122, a drop of approximately 45%).

Methods

To evaluate differences in student achievement, we use a quasi-experimental design with controls for demographic characteristics for students in the treatment and control schools. We compare ESY student outcomes to a control group of students from NKCS elementary schools that maintain traditional calendar years. In addition, we will use nearest-neighbor matching based on factors such as race/ethnicity, English language proficiency, free and reduced lunch status, and grade to compare students in the treatment and control schools. We use difference-in-differences and nearest-neighbor matching methods to evaluate the effect of the intervention on student test score outcomes.

Although NKCS provided student-level MAP assessment data for 2009 to 2017, restrictions were imposed to resolve four main issues: assessment inconsistency, treatment designation switching behavior, missing assessment data, and non-typical grade progression. Details regarding these restrictions are as follows.

First, due to changes in the MAP assessment that extend beyond changes in cut scores and scale, data for the pre-intervention period were restricted to one academic year, 2014-15. Details regarding these changes can be found in the 2015 MAP Technical Manual.

Second, since benefits of increased instructional hours are expected to accumulate over time, students who switched in or out of treatment or control schools during the intervention years are removed from the sample. This group includes students who stayed within the district but switched from one treatment designation in 2016 to a different one in 2017. For example, a student who switches from a treatment school in 2016 to a control school in 2017 is omitted from the sample. This resulted in a loss of 184 observations out of the total 7,632. Half of these observations correspond to student math scores and half student communication arts scores.

Third, due to a limited sample size, it is possible that if many poorly performing students opted out of the intervention and many highly performing students opted into the intervention, the results may be biased. It is important to ensure that results were not driven by students who moved into the district during the intervention or left prior to the intervention, so we dropped students who had missing MAP scores. For Cohort 1 and 2, this means that students must have three years of MAP data, and for Cohort 3, this means that students must have two years of MAP data to be included in the analytic sample. This resulted in a loss of 24 total observations with three being from the communication arts treatment group of 267 students, four being from the communication arts control group of 410 students, 11 being from the communication arts other district group of 3,047 students, and six being from the math other district group of 3047 students.

Fourth, students with non-typical grade progression were dropped from the analytic sample. This includes students who repeated or skipped grades. The rationale for dropping these students is that there is no certain definition for which cohort each student should belong which makes appropriate nearest-neighbor matching unclear. This final restriction resulted in a loss of eight observations from the other district schools group, four students from math and four from communication arts. The effects of these general restrictions on our final student counts may be seen in *Table 4.* The breakdown of the final students counts into treatment groups and cohorts may be seen in *Table 5*.

Results

Difference-in-Differences (DID)

Economists use difference-in-differences methods to examine how a policy affects a treatment and control group differently. Essentially, by comparing the pre-and post- intervention outcomes of the treatment group to pre-and post- intervention outcomes of the control group, researchers can discern whether the outcome, in this case MAP scores, was affected by the ESY intervention. If the treatment has no effect the result of DID estimate will be zero. If the DID estimates are positive, this suggests that the treatment group learned relatively more than the control group after the intervention.

For simplicity, from this point forward we focus our attention on treatment and control group outcomes only. *Table 6* presents pre- and post-intervention means by treatment group, cohorts, and content area. The values of interest are found in bold at the intersection of the *Difference* rows and *Difference* column. These values are the DID estimates for each cohort and content area. Of the six DID estimates, the only estimate statistically significant at at least the 5% level is for cohort 2 in mathematics. This coefficient means that the change in math MAP scores for cohort 2 students is .5655 standard deviations greater for the treatment group than then control group between 2015 and 2017. Regression output with the associated standard deviations corresponding to these DID estimates may be seen in *Appendix B*.

Regression

While looking at means or differences in means is relatively easy to understand, it ignores the possibility that characteristics outside of the intervention itself may influence the outcome of interest. While *Table 6* suggests that the treatment group is moving generally in a positive direction for math, it does not control for other explanatory factors such race, gender, socioeconomic status, English proficiency, summer school attendance, and school level fixed effects.

Table 7 contains the pooled regression results from Difference-in-Difference (DID) models for communication arts and mathematics. The model is specified in equation 1:

$$Y_{cit} = \beta_0 + \beta_1 T_t * X_{it} + \beta_2 T_t + \beta_3 X_i + \beta_4 D_i + \beta_5 S_i + \beta_6 F_i + \varepsilon_i \qquad (1)$$

Y_{cit} is the outcome variable and represents the standardized value of student *i*'s state assessment during time *t* for content area *c* (e.g., communication arts, mathematics). $T_t * X_{it}$ is the variable of interest, the DID Estimator. The DID Estimator is equal to 1 for individuals in the treatment group during an ESY intervention year and 0 otherwise. T_t is an ESY intervention year dummy variable. X_i is a set of dummy variables representing English-lanuguage learner status, free and reduced lunch status, race and ethnicity (e.g., white, black, hispanic), and male. β_0 is the value of the intercept.

These models contain data from all three years for Cohorts 1 and 2 and only the two intervention years for Cohort 3. The numbers corresponding with the variables are estimated coefficients, and the numbers in the parentheses are standard errors which allow us to infer whether the treatment was significantly different from zero.

The variable of interest in this model is the DID Estimator. The DID Estimator tells us whether being in the treatment group in the post-intervention period is associated with improved growth in the outcome variable when controlling for other demographic variables. The DID Estimator is positive for Cohorts 1 (.175, $p = .287$) and 3 (.311, $p = .103$). Although this suggests that treatment is associated with improved reading outcomes for these groups of students compared to their control group peers, the estimated coefficients are not statistically significant. This means that although we have anecdotal evidence, we have no statistical evidence to conclude that reading outcomes have actually improved for Cohorts 1 or 3 as a result of the ESY intervention. The DID Estimator is negative for Cohort 2 (-.104, $p = .498$), but once again, this result is not statistically significant. Free/Reduced Lunch Status, Male, and some racial characteristics are also statistically significant in the MAP communication arts models.

The DID Estimator indicates improved growth in MAP math scores for Cohorts 2 and 3 compared to its corresponding control group. Cohort 2 and 3 treatment students grow .458 ($p = .004$) and .364 ($p = .054$) standard deviations, respectively, more than their control group peers. Math growth was not statistically significant for Cohort 1 (.0589, $p = .743$), meaning that we have statistical evidence that the intervention improves math outcomes for the treatment group in Cohorts 2 and 3 but not Cohort 1. Free/Reduced Lunch Status and some racial characteristics are also statistically significant in the MAP math models.

The increases in the z-scores of Cohorts 2 and 3 for math students exposed to the ESY intervention are very large, statistically significant effects of the ESY intervention. Communication arts and math for Cohort 1 and communication arts for Cohort 3 had positive signs but were not statistically significant. Communication Arts for Cohort 2 had a negative sign and was not statistically significant.

While the basic regression presented in Table 7 controls for a host of explanatory factors, there are additional controls that come easily to mind. Some of these additional explanatory factors were added to the basic regression, and the DID Estimators for each of these regressions are presented in Table 8. The main takeaway from this set of regressions is that even with the addition of these other explanatory variables, the magnitude and level of significance of the statistically significant coefficients is largely unchanged.

Row 1 repeats the DID Estimator from Table 7 above. *Row 2* represents the basic regression with the addition of a dummy variable for whether a student attended summer school. Note that the estimated coefficient for Cohort 3 in reading became statistically significant although the magnitude of the coefficient remained approximately the same. The estimated coefficient for Cohort 3 in math increased in its level of significance from 10% to 5% with a similar magnitude. *Row 3* is similar to *Row 2* in that it attempts to control for summer school attendance for the control group. The difference is that instead of a dummy variable for summer school, it includes a variable that is equal to the percent of summer school that a student is supposed to attend that he or she actually does attend. *Row 4* includes a dummy variable to control for whether a given classroom is run by a new teacher. New teachers are defined as teachers with three or less years of total teaching experience. *Row 5* includes the New Teacher dummy variable as well as the summer attendance rate variable used in *Row 3*. The full output for *Rows 2-5* may be seen in *Appendix C*.

Nearest Neighbor Matching

One problem with the above regressions is that treatment is not randomly assigned. To account for this, we ran a regression using nearest neighbor matching. Nearest neighbor matching uses the characteristics from each person in the treatment group and finds a match for him or her in the control group based on observable characteristics such as race, gender, English proficiency, and free and reduced lunch status. When no observations can be identified that balance characteristics in a similar way, that observation is dropped.

Table 9 shows the results of the nearest neighbor matching regression. Since the regressions in *Table* 7 indicated that *Free / Reduced Lunch Status* held a high degree of explanatory power, nearest neighbors were required to match exactly on that variable for both communication arts and math analyses. This is also true for *Male* in the Communication Arts models. Nearest neighbors were more loosely matched on gender (for math only), pre-intervention MAP scores, English proficiency, and race.

The findings are consistent with the pooled regression results for math seen in *Table 7*. The intervention positively influences math achievement for both Cohorts 2 (.364, $p = .003$) and 3 (.318, $p = .003$). The nearest neighbor results also indicate a positive influence on reading achievement for Cohort 3 (.280, $p = .010$). All of these coefficients are large and statistically significant. The largest of these estimated coefficients is for Cohort 2 Math. This coefficient means that if all students were treated with the ESY intervention, we would expect that MAP math scores would increase by .364 standard deviations on average. The estimated coefficients for Cohort 2 Communication Arts and Cohort 1 Math are positive but not statistically significant. The estimated coefficient for Cohort 1 Communication Arts is negative but again is not statistically significant.

Discussion

Although state assessment data provide a solid foundation upon which to study the relationship between academic achievement and the ESY intervention, there is one noticeable weakness. Students are exposed to the treatment from the time they enter the district until the time they leave. This means that students as young as kindergarten receive the ESY treatment, but we are not able to study their achievement levels until they reach third grade— the first time they participate in the state assessment. District benchmark data is collected from an early age. In reading, this benchmark assessment is the Fountas & Pinnel (F&P). The F&P reading assessment is scored with letters rather than numeric values. There are currently no reliable mappings of the F&P to numeric values. It is possible that younger students benefit differently from the change in policy than older students, and by failing to measure achievement in the earlier grades a potentially important contribution is missed.

Conclusion

In an effort to narrow the achievement gap, the NKCS District extended the school year by 31 instructional days for two of their 21 elementary schools. This study examined whether the ESY policy implemented by NKCS improved achievement of students in the treated schools relative to a control group of schools with students from similar SES backgrounds. Regression analysis indicates that students who were in 4th grade when the ESY program was implemented saw substantial gains in math compared to their control group peers. Including controls for summer school attendance and teacher experience in the basic regression, suggested that in addition to 4th graders seeing relative improvement in math, students who were in 3rd grade during the first year of the ESY program improved relative to control group peers in both communication arts and math. Creating a nearest-neighbor matched sample produced similar results for the same groups of students and content areas.

References

Afterschool Alliance (2014). Race to the Top Applications by State. Accessed June 10, 2014. http://www.afterschoolalliance.org/ARRA%20webpage%20docs/RTT%20State%20Applications%20Excerpts.pdf.

Alexander, Karl L., Doris R. Entwisle, and Linda S. Olson (2001). "Schools, Achievement, and Inequality: A Seasonal Perspective." Educational Evaluation and Policy Analysis 23 no. 2: 171-191.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Patha, and Christopher R. Walters (2012). "Who Benefits from KIPP?" Journal of Policy Analysis and Management 31 no. 4: 837–860. Accessed June 11, 2014. http://dx.doi.org/10.1002/pam.21647.

Bailey, Martha J. and Susan M. Dynarski (2011). "Gains and Gaps: Changing Inequality in U.S. College Entry and Completion." *National Bureau of Economic Research, Working Paper No. 17633.*

Bennett, Pamela R. and Yu Xie. (2003). "Revisiting Racial Differences in College Attendance: The Role of Historically Black Colleges and Universities." *American Sociological Review* 68:567-80.

Coleman, James S., Earnest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Fredrick D. Weinfield, and Robert L. York (1966). *Equality of Educational Opportunity.* Washington, D.C.: Department of Health, Education, and Welfare, Office of Education.

Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsay, and Scott Greathouse (1996). "The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic

Review." Review of Educational Research 66 no. 3: 227-268.

Crow, Karen and Dale Johnson (2010). "A Comparison of Achievement and Attendance in Schools Using Traditional Academic Year Calendars and Year Round Calendars." Journal of Border Educational Research 8 (Spring): 21-30.

Gamoran, Adam, and Daniel A. Long. (2007). "Equality of Educational Opportunity: A 40-Year Retrospective." Pp. 23-47 in R. Teese, S. Lamb, and M. Duru-Bellat (Eds.), International Studies in Educational Inequality, Theory, and Policy. Volume 1: Educational Inequality: Persistence and Change. New York: Springer.

Goldrick-Rab, Sarah and Fabian T. Pfeffer. (2009). "Beyond Access: Explaining Socioeconomic Differences in College Transfer." *Sociology of Education* 82: 101-25.

Graves, Jennifer (2011). "Effects of Year-Round Schooling on Disadvantaged Students and the Distribution of Standardized Test Performance." Economics of Education Review 30 no. 6 (December): 1281-1305.

Jencks, Christopher and Meredith Phillips (1998). *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institution.

Kim, James S. and David M. Quinn (2013). "The Effects of Summer Reading on Low-Income Children's Literacy Achievement from Kindergarten to Grade 8: A Meta-Analysis of Classroom and Home Interventions." Review of Educational Research 83 no. 3: 386-431.

Losen, Daniel J. (2004). ""Graduation Rate Accountability under the No Child Left Behind Act and the Desperate Impact on Students of Color." Pp. 41-56 in *Dropouts in America: Confronting the Graduation Crisis*, edited by Gary Orfield. Cambridge, MA: Harvard Education Press.

McMillan, Bradley J. (2001). "A Statewide Evaluation of Academic Achievement in Year-Round Schools." Journal of Educational Research 95 no. 2: 67-74.

National Center on Time & Learning (2014). Database of Expanded-Time Schools. Accessed June 10, 2014. http://www.timeandlearning.org/db/.

Reardon, Sean F. (2001). "The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations." Pp. 91-115 in *Withier Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane. New York, NY: Russell Sage Foundation.

Redd, Zakia, Christopher Boccanfuso, Karen Walker, Daniel Princiotta, Dylan Knewstub, and Kristin Moore (2012). Bethesda, MD: Child Trends. Accessed June 10, 2014. http://www.childtrends.org/?publications=expanded-learning-time-both-inside-and-outside-of-the-classroom.

Rumberger, Russell W. (2011). *Dropping Out: Why Students Drop Out of High School and What Can Be Done About It.* Cambridge, MA: Harvard University Press.

Wu, Amery D. and Jake E. Stone (2010). "Does Year Round Schooling Affect the Outcome and Growth of California's API Scores?" Journal of Educational Research & Policy Studies 10 no. 1 (Spring): 79-97.

Tables

Table 1: Missouri Assessment Program (MAP) Achievement Levels

MAP Achievement Levels	
Level	**Description**
Level 1: Below Basic	Minimal Understanding
Level 2: Basic	Partial Understanding
Level 3: Proficient	Adequate Understanding
Level 4: Advanced	Thorough Understanding

Table 2: Description of grade-dependent student groups.

	2015	2016	2017	Data Notes
Cohort 1	Grade 4 (pre-treatment)	Grade 5 (treatment)	Grade 6 (post-treatment)	One year of treatment, one year of post-treatment, one year of pre-treatment data
Cohort 2	Grade 3 (pre-treatment)	Grade 4 (treatment)	Grade 5 (treatment)	Two years of treatment, one year of pre-treatment data
Cohort 3	Grade 2 (no data)	Grade 3 (treatment)	Grade 4 (treatment)	Two years of treatment, no pre-treatment data

Table 3: Average MAP Z-scores by cohort and treatment designation, 2016.

	Treatment	Control	Other
		Communication Arts	
Cohort 1	.086	.058	.308
Cohort 2	.052	.153	.334
Cohort 3	.048	.117	.434
		Mathematics	
Cohort 1	.213	.266	.490
Cohort 2	.431	.282	.518
Cohort 3	.34	.171	.553

Table 4: Students omitted from analytic sample due to general restrictions listed above.

	Communications Arts		Mathematics	
	# of Omissions	Sample Size	# of Omissions	Sample Size
Initial Sample		3816		3816
Treatment Switchers	92	3724	92	3724
Missing MAP Scores	18	3706	6	3718
Unusual Grade Progression	4	3702	4	3714

Table 5: Student counts by treatment status, cohort, and content area.

		Communication Arts, Sample Size	Mathematics, Sample Size
Treatment	Cohort 1	100	100
	Cohort 2	82	84
	Cohort 3	82	83
	Subtotal	264	267
Control	Cohort 1	119	120
	Cohort 2	136	137
	Cohort 3	151	153
	Subtotal	406	410
Other	Cohort 1	962	962
	Cohort 2	968	970
	Cohort 3	1102	1105
	Subtotal	3032	3037
Total		3702	3714

Table 6: DID comparing 2015 to 2017 for Cohorts 1 and 2, and 2016 to 2017 for Cohort 3.

Communication Arts

Cohort 1			
	Year = 2017	Year = 2015	
Treatment	-.059	-.003	-.0566
Control	.020	.183	-.1637
Difference	-.0791	-.1861	**.1070**
Cohort 2			
	Year = 2017	Year = 2015	
Treatment	.153	.080	.0730
Control	.203	.109	.0941
Difference	-.0495	-.0284	**-.0210**
Cohort 3			
	Year = 2017	Year = 2016	
Treatment	.281	.048	.2328*
Control	.027	.117	-.0894
Difference	.2533*	-.0689	**.3222**
	Mathematics		
Cohort 1			
	Year = 2017	Year = 2015	
Treatment	-.010	-.032	.0015
Control	.026	.105	-.0786

Difference	-.0362	-.1372	**.1010**
Cohort 2			
	Year = 2017	Year = 2015	
Treatment	.571	-.002	.5732***
Control	.284	.276	.0077
Difference	.2872**	-.2782*	**.5655*****
Cohort 3			
	Year = 2017	Year = 2016	
Treatment	.629	.340	.2885**
Control	.091	.171	-.0803
Difference	.5376***	.1688	**.3688***

***** p<0.01, ** p<0.05, * p<0.1**

Table 7: DID with covariates for Cohorts 1, 2, and 3.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	C1 ELA	C2 ELA	C3 ELA	C1 MATH	C2 MATH	C3 MATH
DID Estimator	0.175	-0.104	0.311	0.0589	0.458***	0.364*
	(0.164)	(0.154)	(0.190)	(0.179)	(0.159)	(0.189)
ESY Year Dummy	-0.125	0.0730	-0.0892	0.159	0.0191	-0.0780
	(0.112)	(0.0973)	(0.113)	(0.122)	(0.101)	(0.112)
Treatment Dummy	-0.391*	-0.00750	0.335*	-0.133	-0.339	0.415**
	(0.209)	(0.291)	(0.191)	(0.228)	(0.302)	(0.190)
Ell	-0.320***	0.0896	0.161	-0.179*	0.138	0.194*
	(0.0906)	(0.0883)	(0.118)	(0.0985)	(0.0906)	(0.116)
FRL	-0.315***	-0.415***	-0.487***	-0.253***	-0.364***	-0.377***
	(0.0809)	(0.0892)	(0.109)	(0.0886)	(0.0929)	(0.109)
Male	-0.224***	-0.160**	-0.350***	-0.0240	0.0732	-0.145
	(0.0695)	(0.0720)	(0.0931)	(0.0760)	(0.0743)	(0.0922)
White	-0.279***	-0.0870	-0.0340	-0.217**	0.0201	-0.0661
	(0.0978)	(0.0981)	(0.138)	(0.107)	(0.102)	(0.135)
Black	-0.368***	-0.218*	-0.0212	-0.381***	-0.275**	-0.133
	(0.114)	(0.116)	(0.149)	(0.124)	(0.118)	(0.146)
Hispanic	-0.302***	-0.347***	0.143	-0.318**	-0.331***	-0.0348
	(0.113)	(0.111)	(0.164)	(0.123)	(0.114)	(0.161)
Constant	0.635	-1.428	0.358*	0.237	0.312	0.250

	(0.423)	(0.929)	(0.183)	(0.464)	(0.968)	(0.182)
Observations	657	654	466	660	663	472
R-squared	0.143	0.150	0.115	0.127	0.176	0.145

***** p<0.01, ** p<0.05, * p<0.1**

Table 8: Estimated DID coefficients associated with various regressions.

VARIABLES	(1) C1 ELA	(2) C2 ELA	(3) C3 ELA	(4) C1 MATH	(5) C2 MATH	(6) C3 MATH
(1) Basic	0.175	-0.104	0.311	0.0589	0.458***	0.364*
	(0.164)	(0.154)	(0.190)	(0.179)	(0.159)	(0.189)
(2) Summer Dummy	0.116	-0.150	0.348*	-0.0904	0.420***	0.384**
+ Basic	(0.188)	(0.156)	(0.189)	(0.205)	(0.161)	(0.189)
(3) Summer Rate	0.128	-0.156	0.326*	-0.0809	0.419***	0.372**
+ Basic	(0.189)	(0.156)	(0.189)	(0.206)	(0.161)	(0.189)
(4) NewTeacher Dummy	0.153	-0.117	0.321*	0.0322	0.430***	0.464**
+ Basic	(0.163)	(0.155)	(0.191)	(0.179)	(0.161)	(0.196)
(5) Summer + Teacher	0.0779	-0.177	0.335*	-0.140	0.384**	0.472**
+ Basic	(0.188)	(0.158)	(0.190)	(0.206)	(0.163)	(0.196)

***** p<0.01, ** p<0.05, * p<0.1**

Table 9: Nearest Neighbor Matching estimates

	Communication Arts			Math		
	Cohort 1	Cohort 2	Cohort 3	Cohort 1	Cohort 2	Cohort 3
Treatment Status	-0.0557	0.0214	0.2802**	0.0762	0.364***	0.318***
	(0.107)	(0.114)	(0.109)	(0.101)	(0.122)	(0.109)
Observations	219	218	233	220	221	236
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1						

Figures

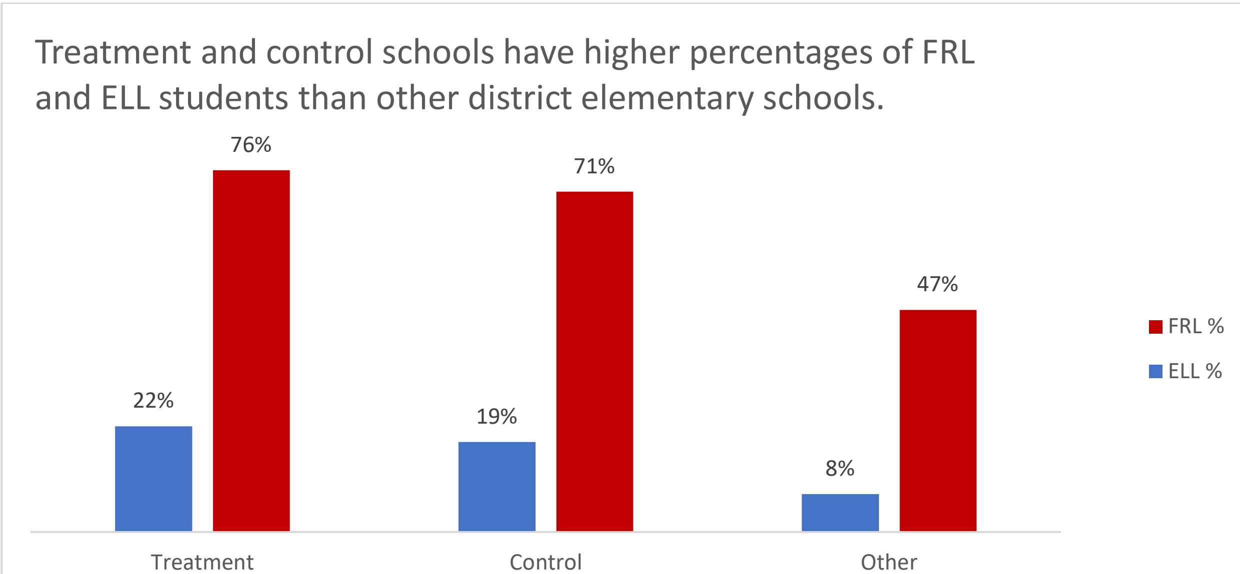

Figure 1: 2015 FRL and ELL percent of total enrollment by treatment status (DESE)

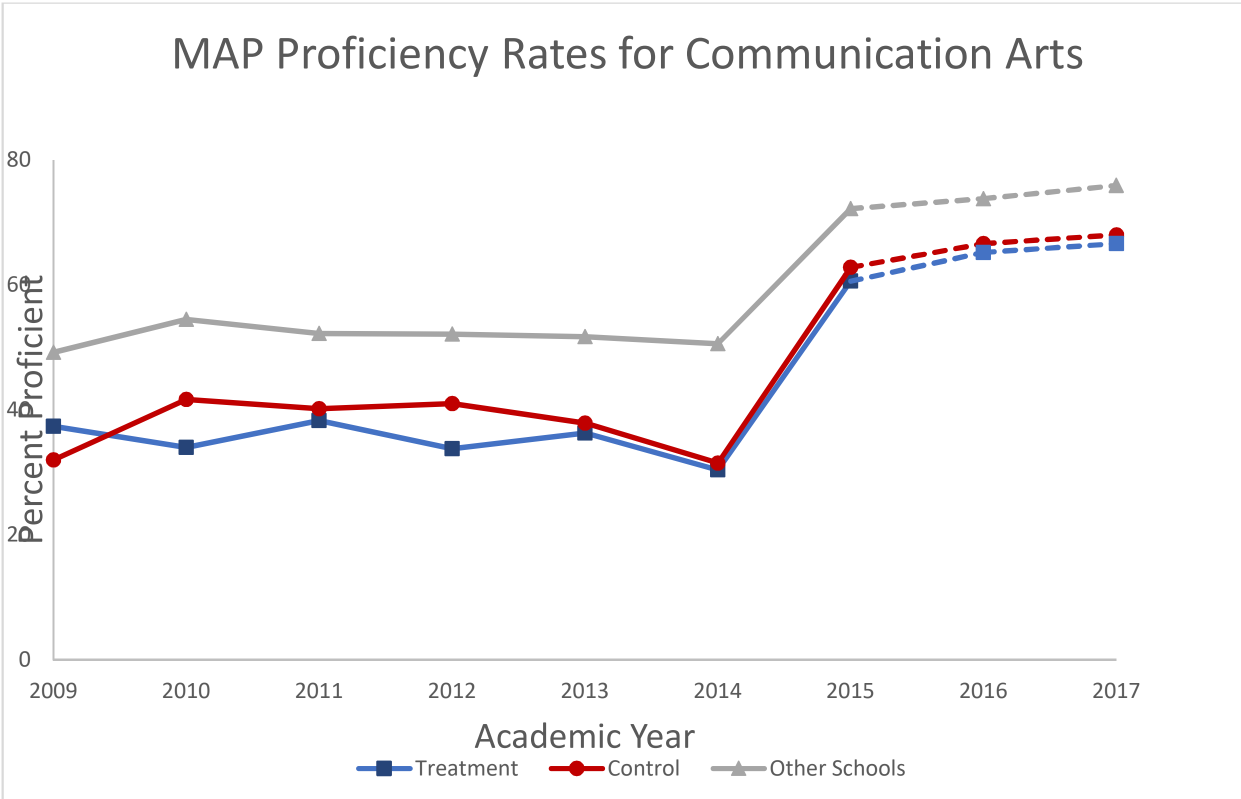

Figure 2: Percentage of third through fifth grade students characterized as Advanced or Proficient in communication arts by treatment status, 2009-2017.

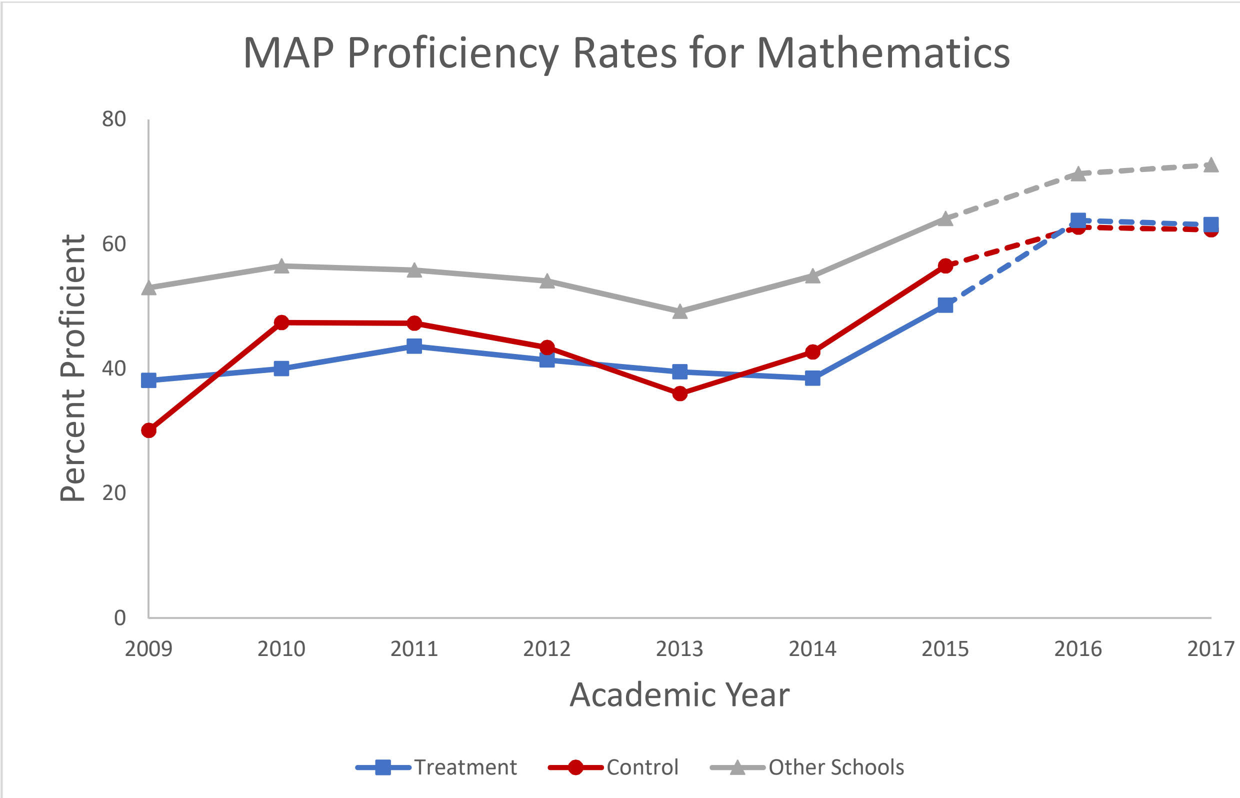

Figure 3: Percentage of third through fifth grade students characterized as Advanced or Proficient in mathematics by treatment status, 2009-2017.

Appendix A

Table 10: NKCSD elementary schools by treatment group

North Kansas City School District Elementary Schools by Treatment Status	
Treatment Schools	Crestview
	Winnwood
Control Schools	Chouteau
	Topping
	West Englewood
Other District Schools	Bell Prairie
	Briarcliff
	Chapel Hill
	Clardy
	Davidson
	Fox Hill
	Gashland
	Gracemor
	Lakewood
	Linden West
	Maplewood
	Meadowbrook

	Nashua
	Northview
	Oakwood Manor
	Ravenwood

Appendix B

Table 11: Regressions underlying the basic DID table.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	C1 ELA	C2 ELA	C3 ELA	C1 MATH	C2 MATH	C3 MATH
DID Estimator	0.107	-0.0210	0.322	0.101	0.565***	0.369*
	(0.181)	(0.187)	(0.200)	(0.187)	(0.198)	(0.199)
ESY Year Dummy	-0.164	0.0941	-0.0894	-0.0786	0.00770	-0.0803
	(0.122)	(0.114)	(0.118)	(0.126)	(0.122)	(0.118)
Treatment Dummy	-0.186	-0.0284	-0.0689	-0.137	-0.278**	0.169
	(0.128)	(0.132)	(0.141)	(0.132)	(0.140)	(0.140)
Constant	0.183**	0.109	0.117	0.105	0.276***	0.171**
	(0.0866)	(0.0809)	(0.0837)	(0.0891)	(0.0865)	(0.0833)
Observations	438	436	466	440	442	472
R-squared	0.009	0.003	0.008	0.003	0.030	0.034

Appendix C

Table 12: Regressions corresponding with Row 2 in Table 8.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	C1 ELA	C2 ELA	C3 ELA	C1 MATH	C2 MATH	C3 MATH
DID Estimator	0.116	-0.150	0.348*	-0.0904	0.420***	0.384**
	(0.188)	(0.156)	(0.189)	(0.205)	(0.161)	(0.189)
ESY Year Dummy	-0.130	0.0551	-0.135	0.145	0.00344	-0.103
	(0.112)	(0.0978)	(0.113)	(0.122)	(0.102)	(0.113)
Treatment Dummy	-0.332	0.0910	0.575***	0.0166	-0.256	0.544***
	(0.228)	(0.296)	(0.205)	(0.249)	(0.309)	(0.204)
ELL	-0.317***	0.0998	0.160	-0.170*	0.147	0.194*
	(0.0907)	(0.0884)	(0.117)	(0.0986)	(0.0908)	(0.116)
FRL	-0.314***	-0.417***	-0.489***	-0.252***	-0.365***	-0.379***
	(0.0809)	(0.0891)	(0.108)	(0.0886)	(0.0928)	(0.109)
Male	-0.223***	-0.156**	-0.355***	-0.0218	0.0770	-0.150
	(0.0696)	(0.0719)	(0.0923)	(0.0760)	(0.0743)	(0.0921)
White	-0.275***	-0.0935	-0.0663	-0.210**	0.0144	-0.0812
	(0.0980)	(0.0981)	(0.137)	(0.107)	(0.102)	(0.135)
Black	-0.367***	-0.215*	-0.00679	-0.380***	-0.274**	-0.121
	(0.114)	(0.116)	(0.148)	(0.124)	(0.118)	(0.146)
Hispanic	-0.301***	-0.343***	0.162	-0.317**	-0.329***	-0.0214

	(0.114)	(0.110)	(0.162)	(0.123)	(0.114)	(0.161)
Summer Dummy	-0.0695	-0.128	-0.351***	-0.175	-0.109	-0.192*
	(0.108)	(0.0776)	(0.114)	(0.117)	(0.0803)	(0.113)
Constant	0.638	-1.539*	0.492***	0.246	0.216	0.324*
	(0.424)	(0.931)	(0.187)	(0.464)	(0.970)	(0.187)
Observations	657	654	466	660	663	472
R-squared	0.143	0.153	0.133	0.130	0.179	0.151

Table 13: Regressions corresponding with Row 3 in Table 8.

VARIABLES	(1) C1 ELA	(2) C2 ELA	(3) C3 ELA	(4) C1 MATH	(5) C2 MATH	(6) C3 MATH
DID Estimator	0.128	-0.156	0.326*	-0.0809	0.419***	0.372**
	(0.189)	(0.156)	(0.189)	(0.206)	(0.161)	(0.189)
ESY Year Dummy	-0.128	0.0529	-0.129	0.148	0.00313	-0.0997
	(0.112)	(0.0977)	(0.113)	(0.122)	(0.102)	(0.113)
Treatment Dummy	-0.344	0.129	0.538***	0.00641	-0.235	0.525**
	(0.229)	(0.299)	(0.207)	(0.250)	(0.311)	(0.206)
ELL	-0.317***	0.104	0.164	-0.168*	0.149	0.195*
	(0.0909)	(0.0884)	(0.118)	(0.0988)	(0.0909)	(0.116)
FRL	-0.315***	-0.416***	-0.495***	-0.254***	-0.364***	-0.382***
	(0.0809)	(0.0890)	(0.109)	(0.0886)	(0.0928)	(0.109)
Male	-0.224***	-0.156**	-0.362***	-0.0234	0.0764	-0.153*
	(0.0696)	(0.0718)	(0.0927)	(0.0760)	(0.0742)	(0.0923)
White	-0.276***	-0.0960	-0.0654	-0.211**	0.0129	-0.0816
	(0.0980)	(0.0980)	(0.137)	(0.107)	(0.102)	(0.136)
Black	-0.367***	-0.216*	-0.0182	-0.378***	-0.275**	-0.129
	(0.114)	(0.116)	(0.148)	(0.124)	(0.118)	(0.146)
Hispanic	-0.301***	-0.349***	0.152	-0.317**	-0.334***	-0.0276
	(0.114)	(0.110)	(0.163)	(0.123)	(0.114)	(0.161)

Summer Rate	-0.0607	-0.162*	-0.307**	-0.180	-0.124	-0.168
	(0.121)	(0.0853)	(0.124)	(0.131)	(0.0882)	(0.124)
Constant	0.637	-1.581*	0.477**	0.245	0.193	0.315*
	(0.424)	(0.931)	(0.189)	(0.464)	(0.971)	(0.188)
Observations	657	654	466	660	663	472
R-squared	0.143	0.154	0.126	0.130	0.179	0.149

Table 14: Regressions corresponding with Row 4 in Table 8.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	C1 ELA	C2 ELA	C3 ELA	C1 MATH	C2 MATH	C3 MATH
DID Estimator	0.153	-0.117	0.321*	0.0322	0.430***	0.464**
	(0.163)	(0.155)	(0.191)	(0.179)	(0.161)	(0.196)
ESY Year Dummy	-0.120	0.0639	-0.0700	0.161	0.00734	-0.126
	(0.110)	(0.0978)	(0.116)	(0.121)	(0.101)	(0.115)
Treatment Dummy	-0.365*	0.0240	0.331*	-0.119	-0.289	0.366*
	(0.207)	(0.292)	(0.191)	(0.228)	(0.303)	(0.191)
ELL	-0.315***	0.0809	0.165	-0.175*	0.125	0.197*
	(0.0898)	(0.0884)	(0.118)	(0.0981)	(0.0907)	(0.116)
FRL	-0.289***	-0.408***	-0.485***	-0.240***	-0.355***	-0.372***
	(0.0803)	(0.0893)	(0.109)	(0.0881)	(0.0930)	(0.109)
Male	-0.206***	-0.150**	-0.349***	-0.00751	0.0824	-0.144
	(0.0690)	(0.0721)	(0.0932)	(0.0757)	(0.0744)	(0.0920)
White	-0.260***	-0.0890	-0.0351	-0.218**	0.0229	-0.0652
	(0.0975)	(0.0981)	(0.138)	(0.106)	(0.102)	(0.135)
Black	-0.381***	-0.188	-0.0218	-0.402***	-0.250**	-0.131
	(0.113)	(0.117)	(0.149)	(0.123)	(0.119)	(0.145)
Hispanic	-0.297***	-0.344***	0.138	-0.323***	-0.330***	-0.0385
	(0.113)	(0.110)	(0.164)	(0.123)	(0.114)	(0.161)
New Teacher	0.238***	-0.0146	-0.151	0.210**	-0.110	-0.255*

	(0.0887)	(0.105)	(0.192)	(0.0941)	(0.107)	(0.139)
Constant	0.605	-1.467	0.346*	0.241	0.256	0.268
	(0.420)	(0.929)	(0.184)	(0.461)	(0.967)	(0.182)
Observations	654	650	466	657	659	472
R-squared	0.152	0.146	0.116	0.135	0.172	0.151

Table 15: Regressions corresponding with Row 5 in Table 8.

VARIABLES	(1) C1 ELA	(2) C2 ELA	(3) C3 ELA	(4) C1 MATH	(5) C2 MATH	(6) C3 MATH
DID Estimator	0.0779	-0.177	0.335*	-0.140	0.384**	0.472**
	(0.188)	(0.158)	(0.190)	(0.206)	(0.163)	(0.196)
ESY Year Dummy	-0.126	0.0433	-0.111	0.148	-0.00968	-0.148
	(0.111)	(0.0982)	(0.116)	(0.121)	(0.102)	(0.116)
Treatment Dummy	-0.290	0.173	0.533**	0.0522	-0.170	0.476**
	(0.228)	(0.301)	(0.207)	(0.250)	(0.313)	(0.208)
ELL	-0.310***	0.0956	0.167	-0.162*	0.137	0.198*
	(0.0901)	(0.0885)	(0.118)	(0.0982)	(0.0909)	(0.116)
FRL	-0.289***	-0.409***	-0.493***	-0.240***	-0.355***	-0.376***
	(0.0804)	(0.0891)	(0.109)	(0.0880)	(0.0929)	(0.109)
Male	-0.206***	-0.146**	-0.361***	-0.00597	0.0863	-0.152*
	(0.0690)	(0.0720)	(0.0928)	(0.0756)	(0.0744)	(0.0921)
White	-0.257***	-0.0975	-0.0662	-0.213**	0.0158	-0.0806
	(0.0977)	(0.0980)	(0.137)	(0.106)	(0.102)	(0.135)
Black	-0.379***	-0.186	-0.0188	-0.401***	-0.249**	-0.127
	(0.113)	(0.116)	(0.148)	(0.123)	(0.118)	(0.145)
Hispanic	-0.296***	-0.347***	0.148	-0.324***	-0.334***	-0.0314
	(0.113)	(0.110)	(0.163)	(0.123)	(0.114)	(0.160)

Summer Rate	-0.0960	-0.170**	-0.305**	-0.219*	-0.136	-0.167
	(0.121)	(0.0861)	(0.125)	(0.131)	(0.0890)	(0.124)
New Teacher	0.241***	-0.0203	-0.138	0.217**	-0.114	-0.255*
	(0.0888)	(0.105)	(0.191)	(0.0940)	(0.107)	(0.139)
Constant	0.608	-1.635*	0.465**	0.252	0.121	0.333*
	(0.420)	(0.931)	(0.189)	(0.460)	(0.970)	(0.188)
Observations	654	650	466	657	659	472
R-squared	0.153	0.151	0.127	0.139	0.175	0.155

Chapter 2: Linking Incongruent Data Sources: A Case Study of a Summer Library Program

This chapter is coauthored with Meghan Ecker-Lyster, Ph.D. and Karin Chang, Ph.D..

Introduction

A wealth of empirical evidence demonstrates that the educational achievement gap is exacerbated by economic inequality. Compared with children from advantaged backgrounds, children from low-income families are less likely to score as proficient on school reading assessments (Jencks & Phillips, 1998; Reardon, 2011), graduate from high school (Losen, 2004; Rumberger, 2011), and attend college (Goldrick-Rab & Pfeffer, 2009; Bailey & Dynarski, 2011). Although there are many underlying causes of income-based disparities, low-income children are more vulnerable to summer learning loss than their wealthier peers (Cooper et al., 1996; Entwisle, Alexander, & Olson, 2000).

To better understand summer learning loss, the "faucet theory" suggests that opportunities to learn and access educational resources are "turned on" for all children during the academic year, and are accessed equally, however the school resources are "turned off" during the summer months when school is not in session (Entwisle et al., 2000). As a result of the faucet being shut off during the summer months, low-income students whose families cannot afford to provide supplemental educational resources and learning opportunities are put at a distinct disadvantage compared to their affluent peers. On average, children from middle-income homes have access to approximately twelve books per child, whereas, children from low-income families have access to about one book per child (Celano & Nueman, 2008).

To combat summer learning loss a number of districts and communities have implemented summer reading programs. Local public libraries are a key player in the delivery of summer reading programs, as this is one community establishment that offers all children, regardless of income level, equal access to learning opportunities (e.g., books). Over 95% of libraries in the U.S. offer a summer reading program (NCES, 2014). Research has found promising evidence regarding the success of public library summer reading programs on enhancing student reading outcomes (e.g., Roman & Fiore, 2010; Los Angeles County, 2001; Shin & Krashen, 2008).

The literature on summer learning loss and summer reading programs address major aspects of the efficacy of these interventions on student outcomes, however the literature leaves important questions around measurement largely unanswered. Many evaluation studies utilize the same standardized reading assessment, such as the Scholastic Reading Inventory (SRI), to measure the impact of library reading programs on student outcomes (Roman & Fiore, 2015). Few studies explicitly explore the methods necessary to successfully leverage extant reading data across multiple school districts, which often contain a myriad of reading assessments.

Study Purpose and Research Questions

A major challenge to evaluating the impact of a library summer reading program (SRP) is accessing the necessary data to adequately measure program success. Many programs do not have the resources nor time to administer a specific reading assessment to participants. A cost-effective alternative is to obtain reading data from the local school districts that send students to the library's reading program. However, school districts choose benchmark assessments individually and there is considerable variability across schools, grade levels and content areas in terms of which benchmark assessment is used. Thus, working with multiple districts, many of whom choose different assessments, poses some methodological challenges.

The purpose of this study was to develop a data-driven, research method that could be used to measure and evaluate the impact of SRP. The primary research question that guided the evaluation study was: Do students who participate in an SRP experience comparable levels of summer learning loss compared to similar students who did not participate in the program?

Data

The SRP that was evaluated in this study was developed by the Mid-Continent Public Library which serves the Kansas City Metropolitan Area (KCMA). Data for this study were derived from two primary sources: the SRP program records and extant academic data from participating school districts. Data for this analysis corresponded to three points in time. Pre-intervention data were denoted *Spring* and came from the spring of academic year 2016-17. Post-intervention data were denoted *Fall* and came from the fall of academic year 2017-18. Intervention data were denoted *Participation* and corresponded to the intervention which occurred during the summer of 2017.

Approximately 11,000 students from 21 public school districts as well as various private and charter schools across the KCMA participated in the 2017 SRP. Data were solicited from 16 of the 21 public school districts, and data sharing agreements were ultimately executed with eight. The 16 districts were chosen primarily due to size. The data sharing agreement specified that *Spring* and *Fall* datasets be submitted as separate rather than aggregate files. This specification necessitated additional time be set aside to create a matched *Spring – Fall* dataset. Commentary regarding the rationale behind this decision and its applicability to areas of high student mobility may be see in *Appendix A*. The distribution of participants across districts and by participation status may be seen in Figure 1.

Spring and *Fall* data were administrative data provided by the eight participating public-school districts. *Spring* and *Fall* data elements consisted of identifiers such as name, birth date, and state identification numbers, demographic information and benchmark reading scores.

Participation data were self-reported by participants and then provided to us by MCPL. The data corresponded to students who participated in the 2017 SRP. *Participation* variables consisted of name, birthdate, school, and grade. Note that *Participation* data did not include numeric identifiers; this created challenges, as discussed later in the paper, when joining the various datasets. A full list of requested elements may be seen in the sample Memorandum of Understanding located in *Appendix B*.

Methods

The analytic sample for this study consisted of a *treatment* group, defined as first through fifth grade students who participated in the 2017 SRP, and a *control* group of similarly aged students who did not participate. *Figure 2* presents the grade distribution of all participants in the SRP as reported by the parents of these participants or by the participants themselves to the MCPL staff. Notice that a substantial percentage of participants did not report grade information. Due to the higher likelihood of inaccuracy in these data, observations were dropped based on district data rather than *Participation*. More specifically, student records were dropped if they could not be linked to district data at all or if once linked to the district data, students appeared to be outside the range of first through fifth grades. Students who were in kindergarten or younger during Spring 2016-17 were dropped from the analytic sample due to lack of assessment measures. Students older than fifth grade were dropped due to small sample sizes.

Linking Data Sources

The decision to match *Spring* with *Participation* data rather than *Fall* with *Participation* was arbitrary, but the decision to match district data to *Participation* before matching district to district was intentional. As evaluators, our primary responsibility was reporting on the efficacy of the program, but due to the annual replication of this study, it was equally important that we assess the quality of data collection related to participation and make recommendations accordingly. *Participation* data, particularly the identification variables such as name and date of birth, were prone to errors, missing information, and as discussed in *Appendix C*, informal variations of names. We wanted to know how much the quality of these data influenced our ability to match *Participation* to district data. By matching *Participation* to *Spring* first, we could determine the percentage of SRP participants that were identifiable in the district data through name-matching. This percentage provides insight into the quality of the *Participation* data and the related collection process.

Standardizing Assessment Scores

The raw data for this evaluation reported student assessment scores across five different reading measures: Lexiles, Fountas & Pinnell (F&P), Curriculum-Based Measure of Oral Reading (RCBM), Rausch Unit Scale (RIT), and STAR Reading (STAR) Scores. Because these assessment scores were based on varying scales, we standardized each to get scores that are comparable across all districts and grades. Student records containing only F&P scores were dropped from the analytic sample. This is because F&P scores are non-numeric and currently there was no agreed upon method by which to quantify these data. While this only affected one of the eight districts, it was the largest of the participating districts and dropped a substantial number of observations. The equation used to standardize the remaining assessments was:

$$Z_i = \frac{Assessment\ Score_i - Sample\ Average}{Sample\ Standard\ Deviation}$$

Sample averages and standard deviations were functions of assessment type and grade. For example, the z-score calculation for a third-grade student taking the RCBM assessment utilized the sample average and standard deviation of the third-grade students across the eight districts that took the RCBM assessment. However, when assessment companies publish norming sample means and standard deviations, these statistics were used in place of our sample-based statistics.

The benefit of standardizing assessment scores was two-fold. First, it allowed us to keep observations of students who took a different assessment in the fall than what he or she took in the spring. This was particularly important because it was not uncommon for districts to change assessments from one academic year to the next, and with a highly mobile student population, it was also likely that students would move from one district to another that took a different assessment. The second benefit of standardizing assessment scores was that scores became comparable across all grades and assessment types. Because assessment scores were then interpreted as the number of standard deviations that a given score was away from the corresponding mean, the downside of standardizing was that it could be difficult for practitioners and lay audiences to understand the findings. The ease in which we could convert z-scores into something meaningful depended on the audience (e.g., parents, practitioners, administrators, etc.) but was necessary because of the inconsistency of assessment types.

Creating the Matched Sample

One difficulty when analyzing treatment effects is a situation that may arise when involvement in the intervention is not randomly assigned. When the mechanism by which students are influenced to enter treatment is also related to the outcome in question, estimating treatment effects without accounting for this will result in biased estimates. For these data, treatment and control groups look dissimilar in terms of observable characteristics which causes some concern. *Figures 3* and *4* illustrate some of these differences

Treatment group students were more likely to be white (61 percent compared to 53 percent), and less likely to be male (47 percent compared to 50 percent), black (17 percent compared to 23 percent), or Hispanic (12 percent compared to 14 percent). Asian and other racial groups were represented similarly in both the treatment and control groups. and *Table 1* presents means by treatment group along with the p-value associated with the null hypothesis that treatment means and control means are equal. Averages by treatment group are presented in *Table 1* along with the p-value associated with the null hypothesis that treatment means and control means are equal.

It also appeared that treatment students were also less likely to receive free or reduced lunch, it was difficult to determine whether that was actually the case. The Community Eligibility Provision (CEP) allows districts that are 40% FRL or higher to designate all students as FRL. One of the districts in our sample used this provision and this may have altered the distribution of this variable across treatment designations. Treatment group students were much more likely to be enrolled in summer school (68 percent compared to 33 percent). Although summer school attendance may have been another important factor in determining the effect of the SRP intervention on student reading outcomes, this variable is missing in a substantial number of cases.

Because of these differences in demographic characteristics and the fact that assignment to the intervention is not random, we opted to create a matched sample. We accomplished this by using the coarsened exact matching algorithm in Stata as described by Blackwell, et al (2009). For this algorithm, characteristics are more broadly defined, which is particularly helpful for continuous variables. Groupings of characteristics are assigned to "bins". Treatment and control observations that are designated to each of these bins based on their characteristics are considered matches. When treatment observations are assigned to a bin for which there are no corresponding control observations, then these treatment observations are dropped and otherwise for control observations. For this study, the coarsened exact matching was based on pre-intervention standardized assessment scores, summer school attendance status, free and reduced lunch status, white, gender. The reason that we included the spring test score as a matching element is that we wanted to ensure that students were starting out a similar level of proficiency. The logical argument behind this idea is that if we look at two students who are similar in all ways except their exposure to the intervention, then the more likely it is that the intervention is the cause for any differences in the outcome variable.

Before the matching algorithm, we had 10,462 control observations and 1,577 treatment observations. The algorithm identified 1,354 control observations without matching treatment observations and eleven treatment with no corresponding controls. Our final sample sizes were 9,108 control and 1,566 treatment observations.

Table 2 contains the averages by matched treatment groups. After matching, averages are statistically equivalent for gender, Hispanic, Asian, other races, and free and reduced lunch status. Summer school attendance status ($p = .0000$), white ($p = .0450$), black ($p = .0159$), and English-language learner status ($p = .0675$) remained statistically dissimilar for the two groups.

Results

Difference-in-Differences

Economists use difference-in-differences methods to examine how a policy or intervention affects a treatment and control group differently. Essentially, by comparing the pre- and post- intervention outcomes of the treatment group to pre-and post- intervention outcomes of the control group, researchers can discern whether the outcome, in this case reading assessment scores, was affected by the SRP intervention. If the treatment has no effect the result of DID estimate will be zero. If the DID estimates are positive, this suggests that the treatment group learned relatively more than the control group after the intervention.

Table 3 presents pre- and post-intervention means by treatment group. The value of interest is found in bold at the intersection of the *Difference* row and *Difference* column. This value is the DID estimate. This coefficient means that the change in standardized reading assessments scores is 5.9 standard deviations greater for the treatment group than the control group between spring and fall. In other words, the treatment group appears to be worse off as a result of the intervention in terms of the amount that they fall between pre- and post-assessments. However, it is worth noting that despite the fact that the treatment group drops by a greater amount, they stay, on average, above the sample mean.

While looking at means or differences in means is relatively easy to understand, it ignores the possibility that characteristics outside of the intervention itself may influence the outcome of interest. While *Table 3* suggests that the treatment group is falling more between spring and fall than the control group, it does not control for other explanatory factors such race, gender, socioeconomic status, and English proficiency.

Table 4 contains the regression results for the DID model presented in *Table 3* as well as two additional columns to represent DID models with covariates.

The estimated equation from the regression found in *Table 4 Column 2* is represented by Equation 1:

$$Y_i = \beta_0 + \beta_1 T_t * X_{it} + \beta_2 T_t + \beta_3 X_i + \beta_4 D_i + \beta_5 S_i + \beta_6 F_i + \varepsilon_i \tag{1}$$

Y_i represents the standardized post-intervention reading assessment score for individual i. The variable of interest, the DID Estimator, is represented by $T_t * X_{it}$. This variable is equal to 1 for treated individuals during the post-intervention period. T_t is a dummy variable representing the post-intervention period. X_i is a set of demographic dummy variables for male, white, black, Hispanic, ELL status, FRL status, and summer school attendance status.

Notice that the sample size dropped by approximately 32 percent between the basic DID regression and the DID regression with covariates (column 2). This is because one of the larger districts, in an effort to ensure privacy for students, had established a policy of refusing to provide information about student FRL status in samples where students were identified. Because student identification is necessary to merge *Spring*, *Fall*, and *Participation* datasets, we were not able to acquire the FRL data for this district. Because of this, we ran the DID regression with covariates excluding the FRL variable. The output for this regression may be seen in *Column 3* of *Table 4*.

The numbers corresponding with the variables in *Table 4* are estimated coefficients, and the numbers in the parentheses are standard errors which allow us to infer whether the treatment was significantly different from zero. This DID Estimator suggests that when observable characteristics are taken into account, treatment group students are even worse off relative to the control group than they appeared in the basic DID analysis. More specifically, the column 2 regression indicates that the intervention is associated with an almost seven standard deviation greater drop in reading scores for the treatment group as compared to the control group. Column 3 estimates a 4.6 standard deviation greater drop between pre- and post-intervention assessments compared to the control group. Both of these results are significant at the 1 percent level.

Nearest Neighbor Matching

Although we parsed our sample using the coarsened exact match algorithm in order to get more comparable observations between the treatment and control groups, we turned to nearest neighbor matching in order to compute a counterfactual for each of the observations that remained in the analytic sample. The nearest-neighbor algorithm estimates the counterfactual for each observation by identifying one or more students who are similar in terms of a collection of designated observable characteristics. More specifically, the comparison group for a single treated student will be made up of one or more students who are untreated but have similar observable characteristics. Treatment and control students matched exactly on the district that they attended in the spring, otherwise, treatment and control students were match based on a weighted function of gender, race, English language learner status, and spring test score.

Once the match was made, the estimate of the counterfactual outcome variable was calculated. The average difference between all students' actual and estimated counterfactual outcomes is called the average treatment effect (ATE).

Table 5 presents our nearest-neighbor estimates of the ATE of the SRP intervention on student reading outcomes. Compared to similar students, SRP participants had better fall reading outcomes following the intervention. The estimated average treatment effect was 1.518 (p =

.116), meaning that reading outcomes would be 1.518 standard deviations higher if all students participated in the SRP compared to when no students participate. The output also indicates that ties-in-distance caused at least one observation to be matched with five other observations. What this means is that the nearest-neighbor matching algorithm identified all equally good observations and averaged the associated outcomes to calculate the counterfactual in an effort to reduce bias. However, this result is not statistically significant.

Columns 2 – 4 include all of the same variables in the basic nearest neighbor analysis but add controls for FRL status and summer school attendance status. Summer school status is added in such a way that students in the treatment group must match exactly on that variable to students in the control group. FRL status is added to the analysis in a way that it becomes part of the weighted function of observable characteristics. *Column 2* includes all of the basic variables plus the FRL status variable and estimates an ATE of 2.253 ($p = .034$). This means that if all students participated in the SRP, post-intervention reading assessment scores would be expected to be 2.253 standard deviations higher on average. This result is significant at the 5 percent level. Including the summer school attendance variable instead of the FRL status variable, *Column 3*, provides a similar result. In this case the ATE is estimated to be 3.552 ($p = .011$) and is once again significant at the 5 percent level. The final specification includes both the summer school attendance as well as the FRL status variables. The ATE is estimated as 5.269 ($p = .001$). This estimate is substantial in size and is statistically significant at the 1 percent level.

Conclusion

Leveraging secondary data sources is a cost-effective approach to evaluating community programs, such as a summer library program. However, complexity is introduced when data is solicited from multiple sources (e.g., school districts, public library), which pose many challenges for linking and summarizing outcomes. The purpose of this study was to highlight a data-driven, research method that can be used to measure and evaluate program outcomes that rely on data from numerous sources. This case study highlights a methodological approach that was used to evaluate the impact of a summer library program on reducing summer learning loss. Results indicated that the SRP was an effective intervention for reducing the impact of summer learning loss for participants.

References

Bailey, M. J., & Dynarski, S. M (2011). Gains and Gaps: Changing Inequality in U.S. College Entry and Completion. *National Bureau of Economic Research, Working Paper No. 17633.*

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *The Stata Journal*, *9*(4), 524-546.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research 66*, 227-268.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2000). Summer learning and home environment. R. D. Kahlenberg (Ed.). *A nation at risk: Preserving public education as an engine for social mobility* (pg. 9-30). New York, NY: Century Foundation Press.

Celano, D., & Neuman, S. B. (2008). When schools close, the knowledge gap grows. *Phi Delta Kappan, 90*, 256-263.

Goldrick-Rab, S., & Pfeffer, F. T. (2009). Beyond Access: Explaining Socioeconomic Differences in College Transfer. *Sociology of Education, 82*, 101-25.

Kim, J. (2004). Summer reading and the ethnic achievement gap. *Journal of Education for Students Placed at Risk, 9*, 169-178.

Kim, J.S., & Quinn, D.M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research, 83*, 386-431.

Jencks, C., & Phillips, M. (1998). *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institution.

Los Angeles County Public Library Foundation (2001). *Evaluation of the public library summer reading: "Books and Beyond . . . Take Me to Your Reader". Final report.* Los Angeles, CA: Public Library Foundation.

Losen, D. J. (2004). Graduation rate accountability under the No Child Left Behind Act and the desperate impact on students of color (pg. 41-56) in *Dropouts in America: Confronting the Graduation Crisis*, edited by Gary Orfield. Cambridge, MA: Harvard Education Press.

National Center for Education Statistics (NCES, 2014). *Services and resources for children and young adults in public libraries*.

Shin, F. H., & Krashen, S. D. (2008). *Summer reading program and evidence*. Boston, MA: Pearson.

Reardon, S. F. (2011). "The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations." Pp. 91-115 in *Withier Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane. New York, NY: Russell Sage Foundation.

Roman, S., & Fiore, C. D. (2010). Do public library summer reading programs close the achievement gap? *Children and Libraries, 8*(3), 27-35.

Rumberger, R. W. (2011). *Dropping Out: Why Students Drop Out of High School and What Can Be Done About It.* Cambridge, MA: Harvard University Press.

Tables

Table 1: Descriptive Statistics by Treatment Status (Full Sample)

	Treated	Untreated	p-value (H_0: $mean_{treated} = mean_{untreated}$)
Male	.474	.503	.0281
White	.613	.528	.0000
Black	.166	.229	.0000
Hispanic	.119	.139	.0277
Asian	.014	.018	.2990
Other race	.088	.086	.8204
ELL	.032	.047	.0096
FRL	.453	.484	.0526
Summer	.676	.331	.0000

Table 2: Descriptive Statistics by Treatment Status (Matched Sample)

	Treated	Untreated	p-value (H_0: $mean_{treated} = mean_{untreated}$)
Male	.474	.494	.1470
White	.612	.585	.0450
Black	.167	.192	.0159
Hispanic	.119	.127	.4079
Asian	.014	.016	.5607
Other race	.088	.080	.2661
ELL	.033	.042	.0675
FRL	.454	.460	.6975
Summer	.676	.375	.0000

Table 3: Difference-in-Differences (DID) estimates by treatment group

	Post-Intervention (Fall)	Pre-Intervention (Spring)	Difference
Treatment	1.953	14.860	-12.9070***
Control	-2.133	4.877	-7.0104***
Difference	4.0858***	9.9823***	**-5.8966*** **

***** p<0.01, ** p<0.05, * p<0.1**

Table 4: DID regression output

	(1)	(2)	(3)
VARIABLES	basic	with cov	with cov
DID Estimator	-5.897***	-6.928***	-4.607**
	(1.973)	(2.600)	(2.118)
Post-Year Dummy	-7.010***	-7.058***	-7.121***
	(0.756)	(0.933)	(0.825)
Treatment Dummy	9.982***	12.23***	11.23***
	(1.395)	(1.858)	(1.520)
Male		-7.768***	-8.042***
		(0.872)	(0.761)
White		-12.03***	-6.981***
		(1.526)	(1.333)
Black		-16.02***	-15.09***
		(1.808)	(1.499)
Hispanic		-5.941***	-4.670***
		(1.856)	(1.648)
ELL		-19.83***	-22.38***
		(2.179)	(1.949)
FRL		-10.15***	
		(0.917)	

Summer School Dummy		-9.998***	-12.42***
		(0.945)	(0.822)
Constant	4.877***	26.75***	20.27***
	(0.534)	(1.665)	(1.389)
Observations	21,348	14,506	18,170
R-squared	0.009	0.046	0.045

***** p<0.01, ** p<0.05, * p<0.1**

Table 5: Nearest-Neighbor ATE Estimates

	(1) Basic	(2) Basic + FRL	(3) Basic + Summer[3]	(4) Basic + FRL + Summer[3]
Average Treatment Effect	1.518	2.253**	3.552**	5.269***
Standard Error	(0.965)	(1.060)	(1.403)	(1.530)
Observations	10,588	8,842	9,085	7,252

***** p<0.01, ** p<0.05, * p<0.1**

[3] Exact match

Figures

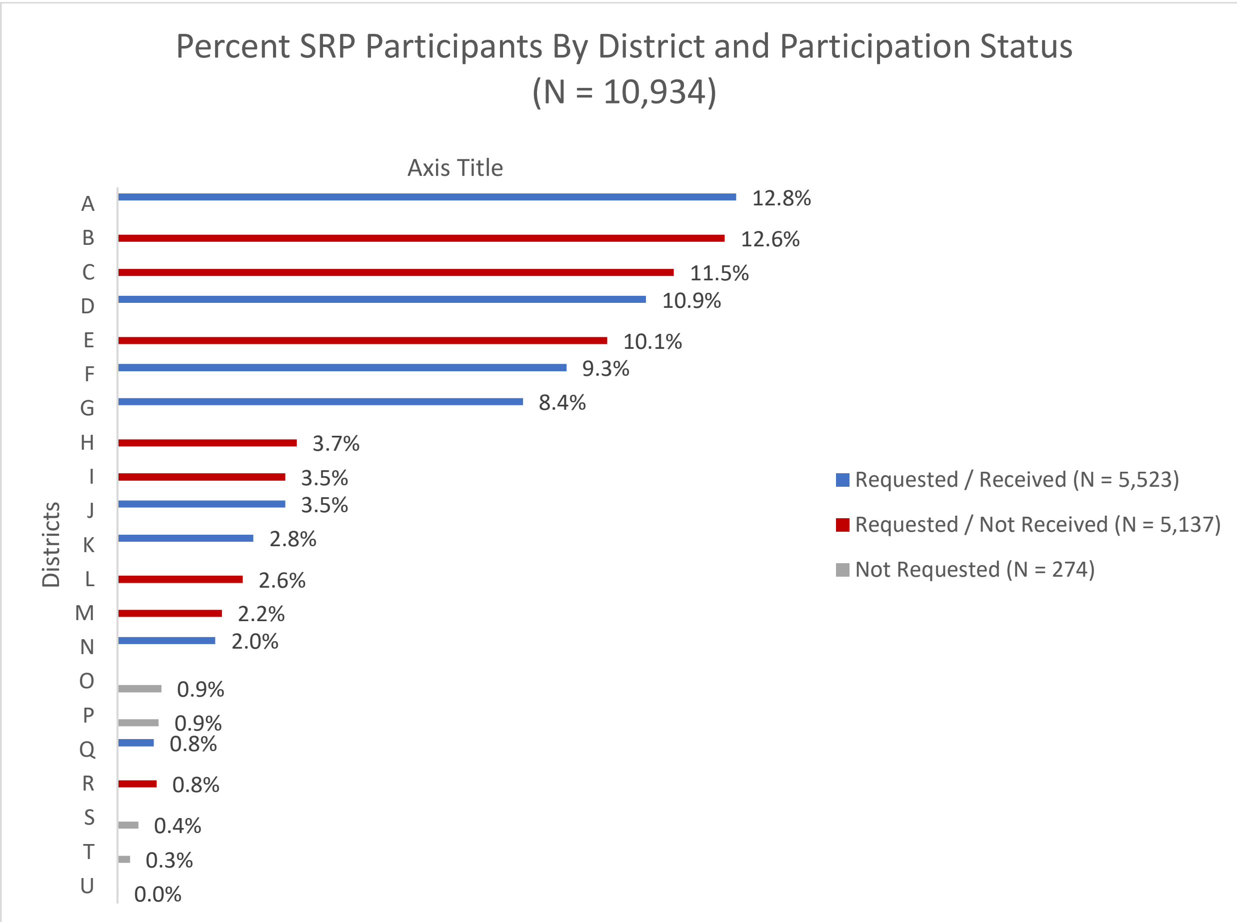

Figure 1: Distribution of SRP participants across school districts

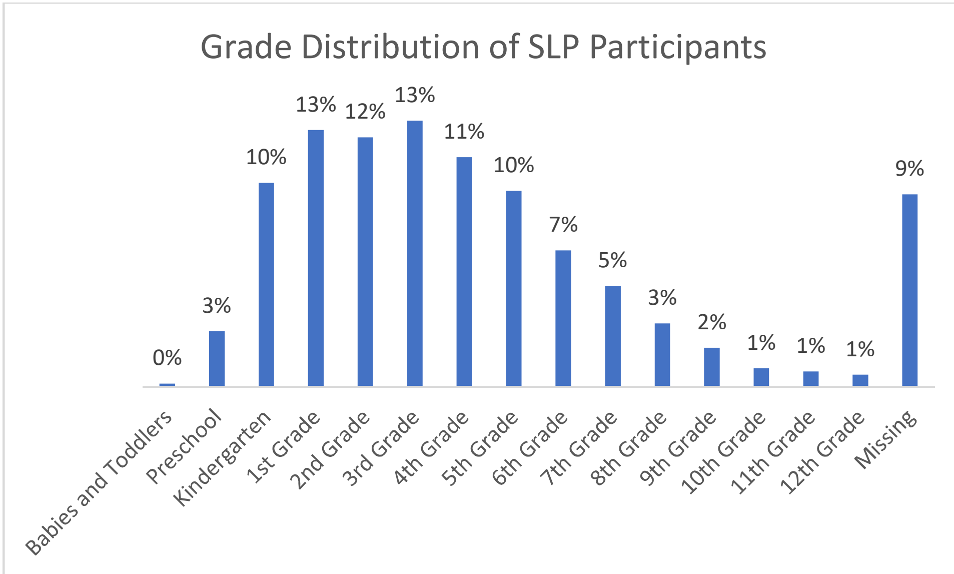

Figure 2: Grade distribution of SRP Participation data

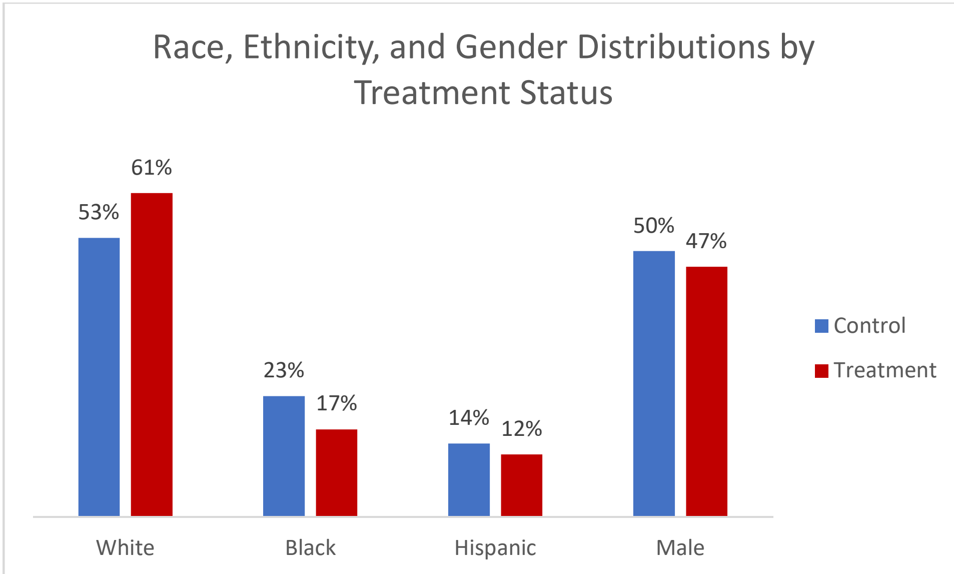

Figure 3: Demographic Characteristics by Treatment Status

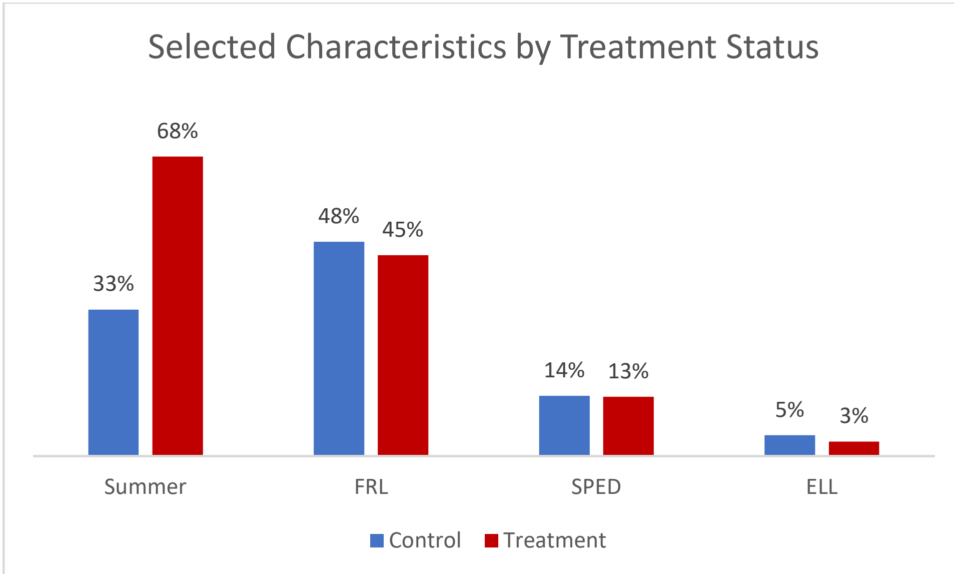

Figure 4: Demographic Characteristics by Treatment Status

Appendix A: Separate versus Aggregate District Data

In the early years of the evaluation process, districts were asked to submit spring and fall data as a single dataset rather than separate files. While not needing to merge spring and fall datasets saved considerable time, there was one major disadvantage. Some of the participating districts have highly mobile student populations, and when submitting the data for analysis, districts often deleted any incomplete student records. In other words, when a student was not enrolled for both spring and fall semesters, districts often dropped the student from the district data submission. Incomplete records are usually dropped during analysis, so under typical circumstances it would not matter whether the district submitted the incomplete student records. Since we work with a number of districts in close proximity to each other, we have an opportunity to identify missing student data across multiple districts. For example, if a student attended District A during the spring and then transferred to District B during the fall, if District B is one of our participating districts then we can maintain that student's record in the analytic sample by simply matching his partial record from District A to his partial record from District B. Although districts would have been willing to resubmit the data to include incomplete student records, we tried to avoid this whenever possible. Because the continued success of this project hinges on district participation, it is important to minimize the effort required by district staff.

Because we received two data files, *Spring* and *Fall*, from each district for the summer 2017 evaluation, we recoded variables for individual districts and then created master spring and master fall datasets which contained all student records from all districts for the corresponding semester. The master *Spring* dataset was matched with MCPL *Participation* data, and then the resulting data were matched with the master *Fall* dataset.

Appendix B: Example Memorandum of Understanding

Memorandum of Understanding between <u>SCHOOL DISTRICT</u>, MID-CONTINENT PUBLIC LIBRARY and the UNIVERSITY OF KANSAS CENTER FOR RESEARCH, INC.

WHEREAS, <u>School District</u>, the Mid-Continent Public Library and the University of Kansas Center for Research, Inc. (KUCR) on behalf of the Kansas City Area Education Research Consortium (KC-AERC) wish to create an independent, non-partisan vehicle of the very highest quality to evaluate the effect of the summer reading program efforts and contribute to basic scholarly research on public schools and educational programming; and

WHEREAS, in order to advance these goals, it is necessary to create a digital data archive consisting of longitudinal data that have been fully cleaned, integrated, and documented; and

WHEREAS, data on student characteristics, student academic performance, and school characteristics are necessary to address the foreseeable research questions of the Consortium and the public it serves; and

WHEREAS, to achieve these purposes, the Consortium will release standardized data to a broad public while protecting the individual-level confidentiality.

Now therefore, the parties agree as follows:

1. <u>School District</u> will appoint a data liaison to coordinate this work (at <u>School District</u>) who will facilitate access to the data, and arrange for the staff resources necessary to create all data files to be provided to KUCR.
2. The Mid-Continent Public Library will appoint a liaison to coordinate this work, facilitate access to the data, and arrange for the staff resources necessary to conduct the

project. The library will support data technician(s) on KUCR staff or as consultants, as necessary and commensurate with the scale and scope of actual data transferred.

3. Once this MOU has been fully executed by all parties, KUCR will provide resources to clean, organize, match, and manage all data files provided for the project. KUCR will design and execute a methodology for analyzing the data.
4. School District will, through a data liaison, or other representative, resolve in a timely fashion through discussions with the Executive Director or other staff of KUCR any questions that arise concerning the content, timing, or other aspects of the data transfer.
5. School District will provide the staff resources needed to assist, in a timely fashion, with the design and documentation of the data, and create or supply the extract files for KUCR from the administrative systems of the School District.
6. School District and Mid-Continent Public Library will provide KUCR with the data needed to sustain the mission of evaluating efforts of the reading program. School District will provide extracts containing the following types of student information for every student enrolled in pre-kindergarten through 12th grade for the school years 2016-2017 and 2017-2018:

Spring, Academic Year 2016-2017	**Fall, Academic Year 2017-2018**
MOSIS (or KIDS) student identification number (scrambled using an agreed-upon algorithm)	MOSIS (or KIDS) student identification number (scrambled using an agreed-upon algorithm)

Student first name	Student first name
Student middle name	Student middle name
Student last name	Student last name
Student date of birth	Student date of birth
Student grade level	Student grade level
Student race	
Student gender	
Student free lunch status	
Student reduced lunch status	
Student special education	
Student English Language Learner (ELL) status	
Student benchmark reading and math assessment scores (e.g., i-Ready, STAR, AIMSweb)	Student benchmark reading and math assessment scores (e.g., i-Ready, STAR, AIMSweb)
Indicate if student attended district summer school in 2017	

7. It is anticipated that modifications to this MOU will be issued for future requested data elements for the following school years: 2018-2019, 2019-2020, and the 2020-2021. Any modification of this MOU shall be in writing and shall be signed by both parties.
8. School District will provide KUCR with the data described above for students in pre-kindergarten through 12th grade attending all schools for which School District maintains data, including elementary, alternative schools, collaborative programs and

special education programs. The feasible beginning date (i.e., historical and longitudinal character) of the data will be determined in consultation with School District.

9. KUCR will preserve the confidentiality of all personally identifiable information about all individual students obtained pursuant to the Memorandum of Understanding in accordance with applicable law, including the Federal Social Security Act, the Family Educational Rights and Privacy Act and any regulations promulgated there under. All studies will be conducted in a manner that does not permit personal identification of parents, teachers and students by persons other than required for research activities undertaken by representatives of KUCR. As such, KUCR will not disclose any such information to any persons except as authorized by law and upon formal approval of School District and will include results in aggregate or in some other non-personally identifiable form. KUCR assures all researchers who are given access to data with individual-level identifiers provided pursuant to the Memorandum of Understanding will have undergone appropriate training.
10. KUCR will subject all research initiated under this Memorandum of Understanding to review and approval by KUCR's Human Research Protection Program, as applicable. KUCR may publish results, analysis, or other information developed as a result of any research based on the data made available under this agreement only in summary or aggregate form, ensuring that no personally-identifiable information is disclosed.
11. KUCR will create a standardized series of data files for broader public release. Standardized data is defined as aggregate school-level data or individual-level data that has been stripped of individual-level identifiers and cannot generate any possible multivariate analysis combining data fields that would yield less than five records per

any data cell. Standardized data files will be reviewed for considerations of accuracy and privacy by School District prior to public release, if a public release is planned.

12. The agreement between School District, the Mid-Continent Public Library, and KUCR is effective as of the date of the last signature and shall continue 60 days after the evaluation agreement authorized by MCPL and KUCR ends. It is anticipated that the evaluation agreement will continue through 02/28/2021, unless terminated earlier by either party. Either party may terminate this agreement provided written notification is received by the other party 30 days prior to the proposed termination date.

By signing below, the official certifies that he or she has the authority to bind the organization to the terms of this Understanding and that the organization has the capability to undertake the commitments in this Understanding.

Appendix C: Difficulties Related to Name-matching

As previously noted, *Participation* data do not include a numeric identifier, so the matching process between *Participation* and *Spring* datasets relied primarily on names. Name-matching these two datasets was a time-intensive endeavor for three main reasons: use of informal name in the *Participation* dataset, inconsistent spelling of names, and missing information. The biggest challenge of the three is the use of informal names. District data provide formal or legal names while *Participation* data are typically more informal (i.e., nicknames). For example, the name "Theodore" in *Spring* might be listed as "Ted", "Teddy" or "Theo" in *Participation*. Although these variations on the name "Theodore" make algorithmic matching difficult, it is still possible to join this student's data from one dataset to the other because these variations are commonplace and widely known. Where this process may become impossible is when informal names are unrelated to the formal name. This is particularly prevalent in districts with high immigrant populations where some students may register under anglicized names. For example, a student legally named "Fan" may decide to sign up for the SRP using the name "Sam". Because these names are phonetically unrelated it may be impossible to match this student's district record with his *Participation* record. It is possible to circumvent this problem by using additional information such as grade, date of birth, and last name, but even with this additional information, students are often unmatchable.

To reduce the amount of time needed to match *Spring* and *Participation* files, we utilized a user-written command in Stata called "matchit" (from Julio Raffo). Matchit measures the distance between two text strings and produces a similarity score for the pairing. If two text strings match exactly then the similarity score is equal to 1, and all other matches result in a similarity score less than 1. Exact matches were automatically kept, and all other potential matches were considered individually. Once *Spring* and *Participation* were matched, the resulting dataset was then matched to *Fall*. This part of the matching process was straightforward in that *Spring* and *Fall* datasets contained numeric identifiers upon which we could connect student records from one semester to another.

Appendix D: Duplicate Student Observations with Mismatched Outcome Data

Once *Spring* and *Participation* are matched, the resulting dataset is matched to *Fall*. The matching process is straightforward in that *Spring* and *Fall* datasets contain numeric identifiers upon which we can connect student records from one semester to another.

One thing to be aware of during this stage of the data preparation is that students, who should appear only once in the fully matched set, sometimes appear more than once. This happens occasionally when students move from one district to another, so student information may exist in both the previous district as well as the current district. For example, a student lives in District A and takes the District A assessment. During that same spring, the student moves into District B, and District B uses a different assessment than District A. District B has the student take their assessment during that same spring. During the fall, the student moves to District C. In the fully matched dataset, the student will have two records. One record will be District A in the Spring with District C in the fall, and the other will be District B in the spring and District C in the fall. We have no reason to conclude that one record is more correct than the other, but each student may only appear in the analytic sample once. Which observation pair do we choose? If we decide to always choose the observation pair with the higher spring score, or likewise with the lower spring score, then we may introduce bias into our estimate of the effect of SRP on student achievement. Although we do not observe this phenomenon very often, it is still important to avoid introducing bias if possible, so we randomly select one observation for students with more than one record.

Chapter 3: The Equal Rights Amendment and Labor Market Outcomes for Women

Introduction

The United States Constitution does not expressly guarantee equal rights based on gender. The Equal Rights Amendment (ERA) was proposed by the National Women's Party in order to remedy this situation. Although the ERA failed its ratification attempt in 1982, the adoption of a completely unrelated amendment has renewed ratification hopes for ERA supporters. This paper examines whether the ERA had an effect on the labor outcomes of women in treated states relative to men and women in control states.

Section 1 of the ERA reads:

> "Equality of rights under the law shall not be denied or abridged by the United States or by any State on account of sex (Mansbridge, 1986, p. 1)."

Background

Although amendments to the U.S. Constitution may be adopted through one of two methods, only one of these methods has ever been used. This particular method requires that a proposed amendment pass in both the House and Senate with at least two-thirds of the votes. If this occurs, the issue is turned over to the states (Archives, August 25, 2013). Typically, states are then given seven years in which to ratify the proposed amendment. An extension can be granted, such as in the case of the ERA, moving this deadline to a total of ten years (Kyvig, 1996). If three-fourths of the states[4] ratify, then the amendment becomes part of the

[4] More specifically, this is ¾ of the states *in the union* at the time of ratification.

Constitution, otherwise the ratification process expires at the deadline (Kyvig, 1996; Mansbridge, 1986).

The ERA's bid to become an amendment followed a typical course. It was introduced in Congress for the first time in 1923. And although it was reintroduced at every subsequent congressional session, it was 1972 before it finally gained approval in both the House and Senate. The future looked promising for ERA supporters as many states rushed to ratify it. In fact, the ratification process in Hawaii began within minutes of the congressional vote and resulted in unanimous approval in both the Hawaiian house and senate (Mansbridge, 1986). By the end of 1972, 22 of the 32 state legislatures in session had ratified the ERA. With six years left to go, ratification of the ERA needed the approval of only 16 more states.

Much to the dismay of ERA activists, support for the amendment soon began to whither. When the seven-year deadline rolled around in 1979, only 35 of the necessary 38 states had passed the ERA. Congress extended the deadline to 1982, but it made no difference. The ERA expired with no additional states passing (Crowley, 2006). [5]

Since the ERA died in 1982, it is reasonable to wonder how it is relevant for discussion today. The reason is twofold. The first can be explained by the ratification story of the 27th Amendment and the second can be explained by the Constitution.

Motivation

The 27th Amendment is sometimes referred to as the "Madison Amendment" because it was one of twelve amendments proposed by James Madison in the late 1700's. Ten of Madison's twelve became our Bill of Rights. As of 1982, only eight states had passed the

[5] Idaho, Nebraska, South Dakota, Tennessee, and Kentucky rescinded the ERA by the end of 1979 but still counted towards the total of thirty-five (Crowley, 2006b).

Madison Amendment. Due to the attention brought to it by a University of Texas student, 203 years after it was first proposed, the Madison Amendment was ratified as our 27th Amendment (Bernstein, 1992).

Many ERA supporters believe that if the Madison Amendment could be passed 200 years after its introduction, it could also happen for the ERA (Denning & Vile, 2000). Whether or not this claim is valid is outside the scope of this paper, but what *does* matter is that people are still advocating the ratification of the ERA. Since the ERA has been around since the 1920's, considerable time and money have been invested in its ratification over the years. In fact, in the last two years, the ERA failed in three states, Arizona, Florida, and Virginia, and passed in two others, Nevada and Illinois, bringing the total numbers of states passing to 37 states. One question that may be asked is whether the ratification process rather than actual ratification induces changes in the states.

Another reason to study the effects of equal rights legislation and the possibility of making that legislation part of the U.S. Constitution is that currently our Constitution does not explicitly prohibit discrimination based on gender. Although some argue that the Equal Protection clause of the 14th Amendment does just that, it was 1971, more than a hundred years after it became an amendment, that the Supreme Court really interpreted it that way (Mansbridge, 1986). When the ERA was introduced in the 1920's it was intended to remedy this constitutional omission.

Although ERA supporters suggest otherwise, it was never really intended to have short run effects. This is because the ERA only guaranteed equality "under the law" and laws at the time the ERA was originally proposed were not overtly discriminatory towards women (Mansbridge, 1986). The ERA would have needed to prohibit discriminatory *behavior* rather than just discriminatory *laws*. Additionally, the ERA applied only to discrimination "by the United States or by any State" rather than to all employers and so therefore excluded non-government employers. The ERA was intended to induce long term change by providing justification for the Supreme Court to rule on future cases with an eye towards gender equality (Mansbridge, 1986).

However, just because the ERA wasn't *intended* to have short run consequences, doesn't mean that it did not. In fact, Crowley (2006) finds that more women living in states that had passed and retained the ERA ran for and were elected to political office compared to women in non-passing or rescinding states. The result is that legislative bodies in ERA states were more likely to surpass a threshold of being at least 15 percent female. This percentage is important because research indicates that it is the threshold beyond which women are more effective and influential in the workplace (Kanter, 1977). Specific to the political arena, Cammisa and Reingold (2004) and Studlar and McAllister (2002) find that the marginal effect of an additional female legislator depends on the number of existing female legislators in a state. Crowley concludes that the ERA campaigns educated women politically, and for many, made politics personal for the first time. Additionally, she suggests that the ideology behind the ERA acted as a symbol of power that prompted women to challenge their traditional roles in society.

One aspect of women's traditional roles in society is their involvement, or lack thereof, in the labor market. This can be examined in two ways. First, we can observe if characteristics of the labor market, in particular those that cannot be controlled by an individual, change for women. Did the probability of employment improve for women? Did they work more hours or have higher salary? Second, we can observe how characteristics that individual women *can* control, such as labor force participation, changed with ERA enactment. Some women have a great level of freedom in making decisions that determine their participation or success in the labor market. In particular, the determinants of married women's labor market decisions may differ from those of unmarried women. Because of this we examine the ERA's effects on labor market outcomes of women by looking at two sets of comparison groups: women compared to men and married women compared to unmarried women.

Equal Rights Legislation

In the 1960's, the U.S. was experiencing many changes especially regarding race and sex discrimination legislation. Two examples of this are the Equal Pay Act of 1963 which mandated equal pay for equal work and Title VII of the Civil Rights Act of 1964 which prohibited discrimination in terms of hiring, firing, and pay (Neumark & Stock, 2006).

In the late 1800's and early 1900's, laws were passed that limited women's employment. These limits involved restricting industries and occupations, times of day, and the maximum number of hours women could work. The strength of the influence that these laws had on women's labor force participation by the mid 1900's is uncertain. But in the mid to late 1960's many states made it clear that these laws violated the terms of Title VII of the Civil Rights Act making it "so that such laws were by and large nonbinding by 1970 (Neumark & Stock, 2006, p. 39)."

In order to identify the causal effects of policy changes, researchers require some variability across the sample. This variability allows for appropriate determination of observations that are affected by the policy change, the treatment group, and observations that are not affected by the policy change, the control group. State level legislation, as opposed to federal legislation, provides this necessary variability and thus the ability to identify causal effects of legislation (Neumark & Stock, 2006).

To avoid the influence of the Equal Pay Act of 1963 and Title VII of the Civil Rights Act of 1964 while gaining the desirable variability for identification purposes, Neumark and Stock (2006) examine the effects of similar state level laws that were enacted prior to 1960. The sex discrimination laws prevalent during this period, known as Equal Pay Laws (EPL), did not tackle the problems of discrimination in hiring and firing, but instead focused on equal pay for equal work. They examine the effects of this legislation on two labor market outcomes: employment and earnings. Their findings suggest that EPLs cause a decrease in the employment rate and an initial decrease in earnings followed by a slow increase over time for women relative to men.

Data

Sample selection

In order to determine the effect of the Equal Rights Amendment on labor market outcomes for women, we use individual level data from the 1970-2012 Current Population Survey (CPS) March Supplement. Data were limited to 2012 in attempt to avoid the influence of current ERA passage. We restrict the sample to individuals 20-65 years of age who live in Standard Metropolitan Statistical Areas (SMSAs). Additionally, we drop those who are in the armed forces, those who work for no pay, self-employed workers, and those for whom state of residence cannot be uniquely determined.[6] This sample selection method applies to all analyses. Additional restrictions that are specific to the analysis of a particular dependent variable are described in the corresponding section.

[6] Prior to 1977, the CPS codes certain states into groups while others are coded individually.

Variables

Dependent

We examine the following labor market outcomes: labor force participation (LFP), employment (EMPL), hours worked (HRSWK), and salary (LNWKRSAL).

LFP_i is a dummy variable equal to 1 if individual i is either employed, not employed and looking for work, or not employed and on layoff. The sample used for the *LFP* analysis is as described above. $EMPL_i$ is a dummy variable equal to 1 if individual i is employed. We used the CPS employment status variable to define employment. The *EMPL* analysis used the additional sample restriction that for all i, individual i is in the labor force (i.e. $LFP_i = 1$). $HRSWK_i$ is equal to the number of hours that individual i reports having worked. The CPS variable for hours worked last week is used for our measure. The *HRSWK* analysis uses the additional sample restriction that for all i, individual i is employed (i.e. $EMPL_i = 1$). $Salary_i$ is defined using the CPS variable for wage income last year. In order to use real rather than nominal values, we adjusted wage income using the personal consumption expenditure deflator (PCE). Next, we divided real wages by the number of weeks worked last year. If weeks worked is equal to zero or missing, then weekly real wages is set to missing. Finally, we add one to weekly real wages and take a log transformation.[7] The resulting unit of measurement is log 2005 dollars. The S*alary* analysis uses the additional sample restriction that for all i, individual i is a fulltime, full year worker (i.e. $Fulltime_i = 1$). $Fulltime_i$ is a dummy variable equal to 1 if individual i worked thirty-five hours per week or more for at least forty-six weeks last year.

[7] For all *i*, real hourly income is greater than or equal to 0. Because log(0)= -∞, we add one to each value of real hourly income so that the resulting log transformation will remain greater than or equal to zero.

Independent Variables

We use various explanatory variables to explore these four labor market outcomes. ERA_i is a dummy variable that is equal to 1 for any individual i who resides in a state during a year for which the ERA is active and 0 otherwise. ERA_i is used to generate our variable of interest, an interaction between the ERA variable, as described above, and a dummy variable indicating that the individual is female. State ERA enactment data was coded based on Crowley (2006).

Other Legislation

Since the goal of these analyses is to identify the causal effect of the ERA on labor market outcomes, it is important to control for other forces that may affect those same outcomes. For example, in the late 1960's and 70's many states passed unilateral divorce laws. Before these laws were passed, states required both spouses to mutually agree to a divorce. Once these laws passed, divorce could be granted based on the desire of only one spouse. While unilateral divorce laws were good in that an individual no longer needed to have a like-minded spouse to get a divorce, it also created an environment of financial uncertainty for married individuals. This may have been particularly worrisome for married women who were less likely to have their own income. Under mutual divorce laws, married women knew that they could depend on having access to their spouses' incomes.

Friedberg (1998) estimates that the divorce rate would have been six percent lower if states had kept mutual divorce laws. As the divorce rate increases, more married women may find themselves facing financial situations that necessitate entry into the labor force. Increased labor force participation among married women may change the composition of the labor force and potentially the likelihood of becoming employed, the number of hours worked, and salary.

For each of the four labor market outcomes, we include a model that includes a time and state varying dummy variable that controls for the passage of unilateral divorce laws.

Unilateral divorce data was coded based on Friedberg (1998). Equal Pay Law data, as previously discussed, was coded based on Neumark and Stock (2006).

A note on the EPL variable: Since EPLs were passed before the earliest survey year in the dataset, the dummy variables are not time varying. Instead, they are effectively dummy variables controlling for a particular group of states. States that passed EPLs may differ in some otherwise unobservable way from states that did not pass EPLs. Seven states passed Equal Pay Laws and not the ERA: Arizona, Arkansas, Florida, Georgia, Missouri, Nevada, and Oklahoma. Eight states passed the ERA but did not pass EPLs: Delaware, Iowa, Kansas, New Mexico, Tennessee, Texas, Vermont, and Wisconsin. Twenty-seven states passed both. Any significant results arising from the inclusion of the EPL variable should not be interpreted as effects generated by EPLs but rather a common trend in states that committed themselves via early legislation to a reduction in sex discrimination.

Other Controls

Another consideration is that what looks like a reaction to the ERA may actually be a result of changes in male employment. In other words, families may choose to substitute the supply of female labor for male labor— meaning women may increase employment while men decrease employment. We account for this by including a control for the male unemployment rate in the third specification for each dependent variable and comparison group.

In addition to the variables described above, the various regressions also include controls for age and aged squared and dummy variables for gender, race, highest level of education, parental status, state and year. Finally, three variables are created to measure each of the first three years leading up to passage of the proposed amendment for each year. The first lead refers to one year prior to the passage of the amendment, the second lead refers to two years prior, and so on.

Table 1 contains weighted means and standard errors for the explanatory variables previously described. State and year dummy variables have been omitted due to space considerations. These descriptive statistics are based on the LFP analytic sample.

Model

Difference-in-Differences (DID) estimation is ideal for assessing the effects of policy changes that influence one group of individuals, the treatment group, but not another, the control group. For this study, our treatment group consisted of females in states with an active ERA law.

One requirement for appropriate DID estimation is that in the absence of the policy change, the treatment and control groups should exhibit similar trends in terms of the outcome under consideration. Several characteristics of the labor market during this time raise some concerns about satisfying this assumption. First, as indicated previously, racial discrimination legislation was passed just prior to the ERA. Second, at certain points in history black women differ from white women in terms of labor force participation rates (Blau, Ferber, & Winkler, 2002). And finally, previous research supports the idea that black women are governed by a different selection mechanism than white women in regard to labor force participation (Neal, 2002). Together this evidence suggests that treatment and control groups may differ in terms of racial composition and those differences may bias our estimates of the ERA legislation. To try to control for this potential problem, in one of the specifications we included an interaction term allowing the effect of the ERA on white women to vary compared to women of other races.

Additionally, we further restrict the sample to observations in Standard Metropolitan Statistical Areas (SMSA) only. This decision was made for two reasons. First, ERA mobilization efforts would have been more concentrated in areas with greatly population density. Exposure to these efforts may have influenced the behavior or decision-making processes of women in SMSAs in ways that rural women were not influenced. Second, survey statistics from the time that the ERA was passed indicate that individuals from urban areas were more pro-ERA as compared to rural residents (Mansbridge, 1986).

Y_{ij} measured the outcomes of labor force participation (Y_{i1}), employment (Y_{i2}), hours worked (Y_{i3}), and salary (Y_{i4}). For each of these dependent variables, we started with the basic model and then provided four additional specifications on that model. For all specifications and models, we control for state (S_i) and year (T_t) fixed effects. Because the error terms may be correlated within a given state, we cluster on state of residence.

The basic specification was of the form represented by Equation 1:

$$Y_{ij} = \beta_0 + \beta_1 (ERA_{it} \, x \, Female_i) + \beta_2 \, Female_i + \beta_3 \, ERA_{it} + \beta_4 \, T_t + \beta_5 \, S_i + u_i \quad (1)$$

The variable of interest is $ERA_{it} \, x \, Female_i$. This variable estimated the effect of ERA on women in states with active ERA legislation. Specifications 2-5 are of the form represented by Equation 2:

$$Y_{ij} = \beta_0 + \beta_1 (ERA_{it} \, x \, Female) + \beta_2 \, Female_i + \beta_3 \, ERA_{it} + \beta_4 \, T_t + \beta_5 \, S_i + \beta_6 X_i + u_i$$

(2)

In specification 2, X_i was a vector of demographic control variables that included age, age-squared, and dummy variables for race, educational attainment, parental status, and marital status. Specification 3 added to 2 by including dummy variables that controlled for Equal Pay Laws, unilateral divorce legislation, and a set control for the years leading up to ERA passage in a given state. Specification 4 included a variable to control for the male unemployment rate. And finally, Specification 5 included an interaction term allowing the effect of ERA to vary for white women.

Results

In *Table 3- Table 6* we present the regression results for our various models. In each table, columns 1-5 represent the specifications for each dependent variable, as described above. Each table and column contain the results of regressions that define men as the control group and women as the treatment group. Thus, these results are intended to gauge the effect of ERA passage on the gender gap. The variable of interest for each of these columns is *Female*ERA*. Column 5 also contains *White*Female*ERA* as a secondary variable of interest.

LFP

Labor Force Participation results may be seen in *Table 2*. In all specifications, the ERA passage variable is associated with statistically significant increases in labor force participation. However, looking at the female interaction term with ERA, regardless of the specification, we found no statistically significant results. The most basic specification, column 1, we found that the estimated effect of ERA passage increased the expected probability of labor force participation of women by approximately .6 ($p = .510$) percentage points relative to men, however, this result is not statistically significant. When we controlled for demographic characteristics of individuals, the estimate rose to .9 ($p = .397$) percentage points but remained insignificant. Adding controls for EPLs, unilateral divorce law changes, and the years leading up to ERA passage, the estimate fell slightly to approximately .8 ($p = .608$, $p = .608$) percentage points and was consistent regardless of whether we included a control for male unemployment rates. Again, these estimates were not significantly different than zero. In Column 5, although the estimated coefficient for the Female-ERA interaction term fell to -1.9 ($p = .266$) percentage points, allowing the effect of ERA passage to vary by race, resulted in an estimate of 3.6 ($p = .005$) percentage point increase for white women. These Female-ERA interaction term is statistically insignificant and the White-Female-ERA interaction is significant at the 1% level. This makes sense if women of non-white racial classifications already had higher levels of labor force participation before ERA passage. Furthermore, for white women, these results agree with the intuition that if wages are expected to increase in the labor market due to equal rights legislation, the opportunity cost of not working increases.

Employment

Employment results may be seen in *Table 3*. In the basic employment specification, Column 1, we find that the ERA increases the likelihood of becoming employed for women relative to men by approximately .3 ($p = .014$) percentage points. Including demographic characteristics of individuals in the sample causes the estimate to drop slightly to .2 ($p = .037$) percentage points. Columns 1 and 2 results are significant at the 5% level. The additional controls in the expanded models, Columns 3-5, cause the results to drop only slightly to as low as .1 percent points, but the statistically significance of the results completely disappears ($p = .167$, $p = .179$, $p = .592$). In opposition to the LFP results, ERA effects on white women do not deviate from that of other racial classifications ($p = .684$).

Hours worked

Hours Worked results may be seen in *Table 4*. Estimates of the ERA effect of women are insignificant in specifications 1-4 ($p = .375$, $p = .398$, $p = .518$, $p = .512$). The only statistically significant result is found in the full specification in Column 5. Although it is only significant at the 10 percent level, the estimates in Column 5 suggest that women in ERA enactment states worked .43 hours ($p = .071$) longer than men per week after the enactment. In this column, the variable that allows ERA effects to vary for women of different races, indicates that white women actually dropped the number of hours they worked relative to men by approximately .70 hours ($p = .013$). Although this result is significant at the 5% level, combining the estimated coefficients yields an estimated decrease of .26 ($p = .409$) hours for white women. This result is not statistically significant. In total, this means that non-white women worked more hours relative to men after the passage of ERA, and white women did not. This result is consistent with white women increasing labor force participation but not being more likely to be employed.

Salary

Salary results may be seen in *Table 5*. The salary regressions imply that the ERA did not reduce the gender gap in pay. Estimates for the first four specifications found no statistical evidence that women's salaries improved as a result of ERA enactment ($p = .411$, $p = .208$, $p = .367$, $p = .364$). The full specification found in Column 5 shows an increase in real hourly salaries for women of 4.4 percent ($p = .002$). This result is statistically significant at the 1 percent level. Unfortunately for white women, the estimate that allows the effects of ERA to vary across races indicates that white women actually see a drop in hourly salary that equates to approximately 3.8 percent ($p = .000$). Together these coefficients sum to a .5 ($p = .621$)

percentage point increase in white women's wages, however the estimate is not statistically significant. This suggests that non-white women saw an increase in salary as a result of ERA passage, but white women did not.

Conclusion

Economic research on the effects of the Equal Rights Amendment is sparse. This study focuses on labor market outcomes such as labor force participation, employment, hours worked, and salary for women in states where the ERA was passed. We find that the Equal Rights Amendment increased the likelihood of labor force participation for white women but did little to increase the likelihood of employment for women generally. Non-white women saw slight increases in the average numbers of hours worked per week relative to men and experienced a decrease in the gender gap in pay.

The expected increase in the likelihood of labor force participation for white women is sizeable after controlling for other policies that may influence labor force participation, namely unilateral divorce laws, Equal Pay Laws, and male unemployment. And although the ERA was not expected to reduce the size of the gender gap, it appears to have done just that, at least for non-white women. Overall, white women were attracted to the labor market, but non-white women reaped the benefits of increased pay and hours of work.

With so little research focusing on the Equal Rights Amendment, there are endless possibilities for future direction. Alternative segments of the population may react to policies such as the Equal Rights Amendment in completely different ways. Some examples might be women with young children or women living in rural communities. Understanding this behavior can be helpful in making better legislative decisions. Also, the ERA's effects on hours worked and salary should be looked at more closely. Women working in male-dominated fields may differ in terms of labor market outcomes when compared to women in female-dominated fields. Enactment of laws such as the ERA can go a long way towards eliminating male to female pay differential, but it is imperative that researchers make a concerted effort to understand likely labor market reactions to such a policy.

References

Archives, N. The Constitutional Amendment Process Retrieved August 25, 2013, from http://www.archives.gov/federal-register/constitution/

Bernstein, R. RB. (1992). Sleeper Wakes: The History and Legacy of the Twenty-Seventh Amendment, The. *Fordham L. Rev., 61*, 497.

Blau, F. D., Ferber, M. A., & Winkler, A. E. (2002). *The economics of women, men, and work* (Fourth ed.): Prentice Hall Upper Saddle River, NJ.

Cammisa, A. M., & Reingold, B. (2004). Women in state legislatures and state legislative research: Beyond sameness and difference. *State Politics & Policy Quarterly, 4*(2), 181-210.

Crowley, J. E. (2006). Moving Beyond Tokenism: Ratification of the Equal Rights Amendment and the Election of Women to State Legislatures*. *Social science quarterly, 87*(3), 519-539.

Denning, B. P., & Vile, J. R. (2000). Necromancing the Equal Rights Amendment. *Const. Comment., 17*, 593.

Friedberg, L. (1998). Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data. *The American Economic Review, 88*(3), 608-627. doi: 10.2307/116852

Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American journal of Sociology*, 965-990.

Kyvig, D. E. (1996). *Explicit and authentic acts: Amending the US Constitution, 1776-1995*: University Press of Kansas.

Mansbridge, J. J. (1986). *Why we lost the ERA*: University of Chicago Press Chicago.

Neal, D. (2002). The measured black-white wage gap among women is too small: National Bureau of Economic Research.

Neumark, D., & Stock, W. A. (2006). The labor market effects of sex and race discrimination laws. *Economic Inquiry, 44*(3), 385-419.

Studlar, D. T., & McAllister, I. (2002). Does a critical mass exist? A comparative analysis of women's legislative representation since 1950. *European Journal of Political Research, 41*(2), 233-253.

Tables

Table 1: Descriptive statistics by gender

	Male	Female
LFP Rate	0.85	0.66
Employment rate	0.87	0.70
Hours Worked	36.27	25.28
Log Weekly Real Salary	6.48	6.00
Fulltime Worker	0.67	0.42
Age	39.43	40.04
Parent	0.39	0.46
Educational Attainment: Less than High School	0.17	0.16
Educational Attainment: High School	0.46	0.49
Educational Attainment: Associate	0.11	0.11
Educational Attainment: Bachelor	0.18	0.17
Educational Attainment: Master	0.09	0.07

Table 2: LFP regression estimates

VARIABLES	(1)	(2)	(3)	(4)	(5)
ERA	0.0469***	0.0272***	0.0342**	0.0345**	0.0335**
	(0.00970)	(0.00902)	(0.0131)	(0.0130)	(0.0127)
Female * ERA	**0.00684**	**0.00905**	**0.00849**	**0.00849**	**-0.0186**
	(0.0103)	**(0.0106)**	**(0.0164)**	**(0.0164)**	**(0.0166)**
Female	-0.196***	-0.191***	-0.253***	-0.253***	-0.163***
	(0.0109)	(0.0112)	(0.0291)	(0.0291)	(0.0299)
White*Female*ERA					**0.0362*****
					(0.0123)
White* Female					-0.116***
					(0.00965)
White		0.0279***	0.0279***	0.0278***	0.0783***
		(0.00510)	(0.00513)	(0.00514)	(0.00529)
Married		-0.0133***	-0.0135***	-0.0136***	-0.0118***
		(0.00359)	(0.00353)	(0.00352)	(0.00362)
Age		0.0427***	0.0427***	0.0427***	0.0427***
		(0.000446)	(0.000451)	(0.000451)	(0.000450)
Age2		-0.000577***	-0.000577***	-0.000577***	-0.000577***
		(4.91e-06)	(4.91e-06)	(4.91e-06)	(4.91e-06)
Parent		-0.0507***	-0.0509***	-0.0509***	-0.0528***
		(0.00339)	(0.00337)	(0.00338)	(0.00342)
Educ < HS		-0.180***	-0.179***	-0.179***	-0.179***
		(0.00476)	(0.00477)	(0.00479)	(0.00465)

Educ = HS		-0.0546***	-0.0543***	-0.0544***	-0.0537***
		(0.00249)	(0.00257)	(0.00255)	(0.00256)
Educ = Assoc		-0.0167***	-0.0168***	-0.0168***	-0.0166***
		(0.00145)	(0.00145)	(0.00145)	(0.00146)
Educ = Mast		0.0426***	0.0429***	0.0430***	0.0429***
		(0.00150)	(0.00152)	(0.00151)	(0.00150)
Div Law			-0.0392**	-0.0388**	-0.0398**
			(0.0182)	(0.0182)	(0.0170)
Female * Div Law			0.0837***	0.0837***	0.0847***
			(0.0309)	(0.0309)	(0.0291)
Lead1			0.0246**	0.0245**	0.0252**
			(0.0106)	(0.0106)	(0.0107)
Lead2			0.0255***	0.0255***	0.0264***
			(0.00869)	(0.00856)	(0.00859)
Lead3			0.00628	0.00402	0.00479
			(0.0105)	(0.0104)	(0.0105)
EPL			-0.0164	-0.0164	-0.0179
			(0.0118)	(0.0117)	(0.0112)
Female * EPL			-0.0108	-0.0107	-0.00789
			(0.0202)	(0.0202)	(0.0193)
Male Unemployment				-0.00181***	-0.00180***
				(0.000374)	(0.000368)
Constant	0.854***	0.223***	0.247***	0.258***	0.218***
	(0.0131)	(0.0134)	(0.0243)	(0.0246)	(0.0235)

Observations	2,761,150	2,737,971	2,737,971	2,737,971	2,737,971
R-squared	0.055	0.140	0.142	0.142	0.143

Table 3: Employment Regression Estimates

VARIABLES	(1)	(2)	(3)	(4)	(5)
ERA	0.00296	-4.06e-05	0.00453**	0.00535***	0.00528***
	(0.00294)	(0.00241)	(0.00197)	(0.00156)	(0.00155)
Female * ERA	0.00315**	0.00271**	0.00238	0.00232	0.00147
	(0.00123)	(0.00127)	(0.00170)	(0.00170)	(0.00273)
Female	-0.0149***	-0.0139***	-0.0193***	-0.0193***	-0.0109***
	(0.000931)	(0.000874)	(0.00303)	(0.00302)	(0.00353)
White*Female*ERA					0.00140
					(0.00341)
White* Female					-0.0108***
					(0.00215)
White		0.0239***	0.0238***	0.0235***	0.0285***
		(0.00177)	(0.00177)	(0.00177)	(0.00195)
Married		0.0150***	0.0149***	0.0147***	0.0148***
		(0.000673)	(0.000669)	(0.000674)	(0.000678)
Age		0.00482***	0.00483***	0.00472***	0.00472***
		(0.000266)	(0.000266)	(0.000268)	(0.000267)
Age2		-5.17e-05***	-5.18e-05***	-5.02e-05***	-5.03e-05***
		(2.99e-06)	(3.00e-06)	(3.02e-06)	(3.01e-06)
Parent		-0.0105***	-0.0106***	-0.0105***	-0.0108***
		(0.000449)	(0.000446)	(0.000459)	(0.000462)
Educ < HS		-0.0414***	-0.0415***	-0.0420***	-0.0420***
		(0.00114)	(0.00115)	(0.00113)	(0.00114)

Educ = HS		-0.0122***	-0.0122***	-0.0125***	-0.0125***
		(0.000690)	(0.000690)	(0.000685)	(0.000685)
Educ = Assoc		-0.00335***	-0.00338***	-0.00340***	-0.00338***
		(0.000422)	(0.000423)	(0.000415)	(0.000418)
Educ = Mast		0.00126**	0.00130***	0.00147***	0.00151***
		(0.000474)	(0.000465)	(0.000475)	(0.000479)
Div Law			-0.00538**	-0.00444**	-0.00452**
			(0.00210)	(0.00192)	(0.00187)
Female * Div Law			0.00657**	0.00657**	0.00678**
			(0.00320)	(0.00319)	(0.00302)
Lead1			0.00542*	0.00514*	0.00515*
			(0.00290)	(0.00275)	(0.00277)
Lead2			0.00327	0.00333	0.00338
			(0.00251)	(0.00231)	(0.00233)
Lead3			0.00844***	0.00303	0.00301
			(0.00251)	(0.00206)	(0.00207)
EPL			-0.00308**	-0.00281**	-0.00290**
			(0.00132)	(0.00112)	(0.00109)
Female * EPL			-4.02e-06	5.02e-05	0.000275
			(0.00223)	(0.00223)	(0.00222)
Male Unemployment				-0.00431***	-0.00431***
				(0.000245)	(0.000245)
Constant	0.975***	0.859***	0.862***	0.890***	0.886***
	(0.00432)	(0.00586)	(0.00565)	(0.00590)	(0.00609)

Observations	2,074,418	2,056,122	2,056,122	2,056,122	2,056,122
R-squared	0.002	0.017	0.017	0.018	0.019

Table 4: Hours of Work Regression Estimates

VARIABLES	(1)	(2)	(3)	(4)	(5)
ERA	1.097**	0.699*	-0.216	-0.186	-0.209
	(0.433)	(0.413)	(0.338)	(0.353)	(0.355)
Female * ERA	-0.258	-0.242	-0.198	-0.201	0.433*
	(0.288)	(0.284)	(0.304)	(0.304)	(0.235)
Female	-4.987***	-4.968***	-5.374***	-5.373***	-3.533***
	(0.165)	(0.169)	(0.362)	(0.362)	(0.308)
White*Female*ERA					-0.695**
					(0.271)
White* Female					-2.376***
					(0.195)
White		0.494***	0.507***	0.499***	1.892***
		(0.103)	(0.102)	(0.102)	(0.0839)
Married		0.340***	0.346***	0.339***	0.374***
		(0.0427)	(0.0435)	(0.0434)	(0.0390)
Age		0.899***	0.897***	0.893***	0.893***
		(0.0105)	(0.0104)	(0.0104)	(0.0105)
Age2		-0.0105***	-0.0105***	-0.0104***	-0.0104***
		(0.000110)	(0.000109)	(0.000109)	(0.000111)
Parent		-0.938***	-0.936***	-0.935***	-0.990***
		(0.0795)	(0.0788)	(0.0789)	(0.0836)
Educ < HS		-2.423***	-2.405***	-2.425***	-2.414***
		(0.115)	(0.114)	(0.114)	(0.113)

Educ = HS		-1.368***	-1.364***	-1.378***	-1.357***
		(0.0426)	(0.0423)	(0.0420)	(0.0415)
Educ = Assoc		-1.311***	-1.305***	-1.305***	-1.298***
		(0.0422)	(0.0413)	(0.0413)	(0.0416)
Educ = Mast		2.011***	2.012***	2.018***	2.032***
		(0.0792)	(0.0792)	(0.0787)	(0.0790)
Div Law			-0.173	-0.139	-0.157
			(0.159)	(0.164)	(0.159)
Female * Div Law			0.592	0.592	0.664*
			(0.392)	(0.392)	(0.333)
Lead1			-0.813**	-0.824**	-0.838**
			(0.357)	(0.365)	(0.366)
Lead2			-0.331	-0.329	-0.339
			(0.401)	(0.417)	(0.418)
Lead3			-2.492***	-2.686***	-2.707***
			(0.498)	(0.540)	(0.544)
EPL			0.533**	0.544**	0.538**
			(0.208)	(0.216)	(0.222)
Female * EPL			-0.178	-0.176	-0.134
			(0.279)	(0.279)	(0.260)
Male Unemployment				-0.156***	-0.156***
				(0.0110)	(0.0110)
Constant	41.85***	25.01***	25.64***	26.66***	25.54***
	(0.664)	(0.663)	(0.441)	(0.452)	(0.479)

Observations	2,004,765	1,987,143	1,987,143	1,987,143	1,987,143
R-squared	0.068	0.107	0.107	0.107	0.110

Table 5: Salary Regression Estimates

VARIABLES	(1)	(2)	(3)	(4)	(5)
ERA	0.0749***	0.0276**	0.0382***	0.0375**	0.0368**
	(0.0147)	(0.0115)	(0.0142)	(0.0145)	(0.0143)
Female * ERA	0.0150	0.0168	0.0112	0.0112	0.0438***
	(0.0181)	(0.0132)	(0.0123)	(0.0123)	(0.0137)
Female	-0.272***	-0.258***	-0.286***	-0.286***	-0.229***
	(0.0132)	(0.00994)	(0.0228)	(0.0228)	(0.0253)
White*Female*ERA					-0.0380***
					(0.00850)
White* Female					-0.0741***
					(0.0110)
White		0.0933***	0.0933***	0.0935***	0.139***
		(0.00771)	(0.00769)	(0.00769)	(0.0107)
Married		0.107***	0.107***	0.107***	0.108***
		(0.00290)	(0.00290)	(0.00290)	(0.00303)
Age		0.0519***	0.0519***	0.0520***	0.0520***
		(0.000799)	(0.000802)	(0.000804)	(0.000812)
Age2		-0.000504***	-0.000504***	-0.000506***	-0.000506***
		(8.29e-06)	(8.34e-06)	(8.36e-06)	(8.39e-06)
Parent		0.0331***	0.0330***	0.0330***	0.0308***
		(0.00441)	(0.00440)	(0.00440)	(0.00409)
Educ < HS		-0.636***	-0.637***	-0.636***	-0.635***
		(0.0244)	(0.0244)	(0.0244)	(0.0245)

Educ = HS		-0.348***	-0.348***	-0.348***	-0.347***
		(0.00504)	(0.00503)	(0.00503)	(0.00499)
Educ = Assoc		-0.205***	-0.205***	-0.205***	-0.205***
		(0.00396)	(0.00393)	(0.00392)	(0.00389)
Educ = Mast		0.162***	0.162***	0.161***	0.162***
		(0.00467)	(0.00466)	(0.00464)	(0.00470)
Div Law			-0.0275	-0.0283	-0.0289
			(0.0218)	(0.0217)	(0.0212)
Female * Div Law			0.0245	0.0245	0.0278
			(0.0263)	(0.0263)	(0.0242)
Lead1			0.0286	0.0288	0.0281
			(0.0188)	(0.0189)	(0.0188)
Lead2			0.0319**	0.0318**	0.0310**
			(0.0138)	(0.0138)	(0.0137)
Lead3			-0.0176	-0.0128	-0.0139
			(0.0205)	(0.0205)	(0.0203)
EPL			-0.121***	-0.121***	-0.121***
			(0.0106)	(0.0108)	(0.0105)
Female * EPL			0.0157	0.0157	0.0163*
			(0.00989)	(0.00989)	(0.00945)
Male Unemployment				0.00385***	0.00387***
				(0.000702)	(0.000697)
Constant	2.908***	1.791***	1.819***	1.794***	1.758***
	(0.0108)	(0.0175)	(0.0224)	(0.0227)	(0.0225)

Observations	1,472,263	1,461,201	1,461,201	1,461,201	1,461,201
R-squared	0.068	0.316	0.316	0.316	0.317