ABSTRACT

LIU, SIYAN. Dissertation on Estimating School Choice Effects Using Centralized School Assignments. (Under the direction of Melinda Morrill and Thayer Morrill.)

This dissertation provides empirical evidence on estimating the treatment effects of school choice programs using quasi-experimental data. Chapter 1 implements the new econometric method proposed by Abdulkadiroğlu et al. (2017a) in the context of the magnet school program in Wake County, North Carolina. Chapter 2 gives an in-depth analysis of the assignment propensity scores used in Chapter 1. Using simulation data, this chapter proposes two measures for the "ideal" propensity score distribution for the identification of treatment effects.

This dissertation contributes to the recent line of literature on estimating school choice effects using data from centralized assignments. The magnet program in Wake County, North Carolina, is leveraged to address school overcrowding issues by attracting students from overcrowded traditional schools to magnet schools. Chapter 1 evaluates how the magnet program helps students from overcrowded base schools and accounts for differential treatment effects for two student subgroups that vary by their base school overcrowding status. In doing so, this chapter shows an additional benefit of the magnet program of alleviating overcrowding without harming student outcomes.

School overcrowding is a pressing issue faced by many large urban school districts. Chapter 1 evaluates the role of an existing school choice program in alleviating overcrowding in Wake County, NC. The Wake County magnet school program uses centralized assignment to reallocate students with policies to attract students from affluent areas to undersubscribed schools in areas with a larger proportion of economically disadvantaged students. Wake County leverages this system to address school overcrowding at traditional schools through prioritizing applicants of students from overcrowded schools. This paper fully exploits the random variation in the offer of magnet seats at all magnet schools through conditioning on assignment propensity scores and causally identifies the effect of magnet attendance by overcrowding status of origin schools following the method in Abdulkadiroğlu et al. (2017a). Results suggest no evidence of differential gains in maths and literacy skills from magnet attendance by students from overcrowded base schools. However, students from a high-performance overcrowded base school experience higher magnet effects on reduced absenteeism compared to students from a high-performance non-overcrowded base school.

Chapter 2 closely examines the method used in Chapter 1 by Abdulkadiroğlu et al. (2017a). Their approach fully exploits the random variation created by lotteries through conditioning on assignment propensity scores, or the probability of receiving an offer from schools. Chapter 2 studies how characteristics of an assignment process contribute to differences in values and distributions of applicants' propensity scores. I develop measures for the "ideal" propensity score distributions for practitioners using the propensity score conditioning method. I then examine the relationship between these measures and causal identification through simulated data on student outcomes. This chapter gives intuition behind the random variation captured by propensity scores and evaluates the precision of treatment

effect estimates given a propensity score distribution.

Chapter 3 showcases my other line of research on retirement. This chapter studies the income replacement rates of North Carolina public employees as they transition to retirement using administrative records merged with survey data. Results are compared with a nationally representative dataset based on the Health and Retirement Study (HRS). Over half of the older workers have decreased household income transitioning into retirement. Factors contributing to post-retirement income declines include pre-retirement salaries and the lack of higher education. More than 20 percent of our sample experience financial distress measured by having credit card debt, despite having relatively high pension benefit replacement rates. Retirees with income decreases transitioning into retirement are more likely to be financially distressed.

Dissertation on Estimating School Choice Effects
Using Centralized School Assignments

by
Siyan Liu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2021

APPROVED BY:

Robert Clark	Umut Dur
Melinda Morrill Co-chair of Advisory Committee	Thayer Morrill Co-chair of Advisory Committee

DEDICATION

To my parents.

BIOGRAPHY

Siyan Liu was born in Beijing, China. She attended the University of Hong Kong and received a Bachelor of Business Administration with a major in accounting and finance. She also earned a Master of Arts in Economics at Vanderbilt University before entering the Ph.D. program in Economics at North Carolina State University. Her research interests include school choice, education policies, retirement, and financial well-being.

ACKNOWLEDGEMENTS

I would like to express my greatest appreciation to my advisors, Melinda Morrill and Thayer Morrill, for their continued support and guidance. I would not have made it through without their encouragement. I am so lucky to have worked with both of them. Melinda is my role model for her motivation and discipline in research. Her comments on my writing and presentation are always spot-on. Thayer's passion for research has been inspirational. I appreciate that he has always encouraged me to think outside of the box as a researcher.

I would like to thank my other committee members, Robert Clark and Umut Dur. Robert Clark has taught me so much about doing research and communicating with stakeholders - it was my honor working with him as his research assistant and supervisee in grant projects. I am deeply grateful to Umut Dur for his valuable suggestions to improve this dissertation. I would also like to thank Robert Hammond, from whom I first learned about the topic of school choice and how to conduct centralized assignments. I have also benefited from engaging discussions with other faculty members and fellow graduate students at NC State University.

Throughout the years in graduate school, my parents have always believed in me and offered moral support. Thank you, Mom, for showering me with love. I would not have kept my sanity if not for my dearest friends, Naihui and Lu, who reminded me to practice self-love in graduate school. Lastly, I thank my partner, Xun, for being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

DOES SCHOOL CHOICE HELP STUDENTS IN OVERCROWDED SCHOOLS? EVIDENCE FROM A CENTRALIZED SCHOOL ASSIGNMENT

1.1 Introduction

Rapid population growth and budget shortfalls have led to overcrowding in many urban school districts (**?**NCES, 2000, 2007). As existing facilities fail to keep up with student enrollment growth, school districts may choose to leverage existing school choice programs to solve overcrowding. The objective of maximizing existing facility utilization can be embedded in the assignment design, attracting students to less-crowded schools. This paper studies how students from overcrowded schools benefit in cognitive and non-cognitive outcomes from the magnet school programs in Wake County, North Carolina, and proposes a novel approach to estimating overcrowding effects.

Wake County is one of the fastest-growing urban areas in the nation and faces the challenge of keeping pace with student population growth while delivering quality learning environments. NCPS (2016) identifies Wake County as having the greatest need in North Carolina for new school construc-

[0]I would like to express my great appreciation to my advisors Melinda Morrill and Thayer Morrill for their continued support and guidance. I also thank Robert Clark, Umut Dur and Robert Hammond for their valuable comments and suggestions. Access to confidential data was given by the department of Data, Research, and Accountability at the Wake County Public School System (WCPSS).

tion between 2015 and 2020.[1] In Wake County, grade-entry students are assigned to a "base school" according to their home addresses. The Wake County magnet program allows students to apply to receive magnet seats in an annual centralized magnet reassignment. Apart from promoting diversity and boosting student performance, another essential purpose of the magnet program is to alleviate school overcrowding with school choice by "maximizing use of school facilities" (WCPSS, 2019b). Given this policy objective, this paper focuses on school overcrowding in traditional schools and not in magnet schools. The design of the magnet schools' priority structure over students reflects this intention with priorities given to students applying from an overcrowded base school, thus increasing their chances of a successful reassignment to magnet schools.

The empirical challenge in estimating the effect of school overcrowding on student achievement is that students sort into overcrowded schools. High- or low-performing students may enroll at overcrowded schools, depending on the reason for overcrowding. Parents may choose to move to areas where neighborhood schools have desirable characteristics, raising the demand for school seats and resulting in overcrowding. It is also possible that schools in areas with a higher proportion of low socioeconomic status students have more difficulty obtaining resources to build and maintain school facilities to accommodate student growth.[2]

In the context of Wake County's magnet program, this paper evaluates whether students experience achievement gains from "escaping" overcrowded traditional schools and attending magnet schools. I separate magnet applicants from traditional schools into two groups by their base school overcrowding status. The group of applicants with an overcrowded traditional base school likely "fall back" to their overcrowded base school if they do not receive a magnet offer, while others "fall back" to a non-overcrowded traditional base school.

This paper adds to the few studies on school overcrowding effects (**?**Rivera-Batiz and Marti, 1995; Shirley, 2017). Two common solutions to school overcrowding have been examined by the literature—mobile classrooms and multi-track year-round school calendars. I present an alternative solution that utilizes an existing school-choice program to attract students away from overcrowded schools. I also propose a new approach to estimate the effect of school overcrowding. Although school overcrowding status is likely non-random for students, centralized assignment allows for random variation in magnet school offers. Taking advantage of the random variation in centralized school assignments, I provide causally identified estimates of the effect of magnet attendance by base school overcrowding status following the econometric method proposed by Abdulkadiroğlu et al. (2017a). I then examine the differential gain from attending a magnet school by base school overcrowding status as a proxy of overcrowding effects among applicants to the magnet program. While I do not find evidence of overcrowding effects on math and literacy skills, the results show a large and statistically significant overcrowding effect on absenteeism. Compared to students with a high-performance, non-overcrowded

[1] In 2015, 59 of the 107 elementary schools in Wake County had more enrolled students than the capacity of their permanent buildings and temporary classrooms combined.

[2] Prior to 2010, Wake County implemented large-scale reassignment policies with the purpose of reducing school overcrowding in areas experiencing high population growth and maintaining socioeconomic diversity. Students from both high- and low-income neighborhoods were reported to be reassigned and bused to attend less overcrowded schools, suggesting both types of overcrowded schools may exist in Wake County (Hoxby and Weingarth, 2005; Parcel and Taylor, 2015).

base school, students with a high-performance, overcrowded base school experience a more than 40% reduction in the number of absences from attending magnet schools.

This paper contributes to a recent line of research that identifies magnet school effects using the methodology by Abdulkadiroğlu et al. (2017a), which fully exploits the conditional random variation in assignment outcomes (Abebe et al., 2019; Abdulkadiroğlu et al., 2020; Dur et al., 2020; Winters, 2020). Abdulkadiroğlu et al.'s (2017) approach extracts an assignment probability or "propensity score" of each applicant based on their submitted preferences and priorities and the assignment mechanism applied for the centralized assignment. They show that conditioning on the propensity score alone is sufficient to ensure the independence of school offers with respect to student characteristics, making them valid instruments for school enrollment (Rosenbaum and Rubin, 1983). This approach fully utilizes the random variation in the assignment and allows for identifying school choice effects from a larger quasi-experimental sample. Dur et al. (2020) adopts Abdulkadiroğlu et al.'s (2017) propensity score conditioning method to evaluate the impact of magnet schools in the Wake County Public School System. This paper adopts the approach in Dur et al. (2020) to identify magnet effects for two subgroups of magnet applicants by their base school overcrowding status.

The remainder of this paper is organized as follows. Section 2 provides background information regarding the literature on school overcrowding and Wake County schools. Section 3 provides details on the Wake County centralized magnet school assignment, the assignment algorithm used in Wake County, and the assignment propensity scores. Section 4 explains the empirical methodology used in this paper, with data descriptions discussed in Section 5. Section 6 describes school overcrowding effects on student outcomes, and Section 7 concludes.

1.2 Background on Overcrowding and Wake County Assignment

1.2.1 Literature Review

1.2.1.1 School Overcrowding

There is little quantitative evidence on the direct effects of school overcrowding on student outcomes. Rivera-Batiz and Marti (1995) use school-level regression analysis and finds that overcrowding in New York City public schools has a negative effect on student achievement in schools with students from low socioeconomic status. For students with high socioeconomic status, school overcrowding is positively correlated with student achievement, which they explain is a result of better schools attracting more students. The positive correlation can be evidence of unobserved student characteristics, which likely correlate with both enrollment in an overcrowded school and academic outcomes. Shirley (2017) uses the fixed effects model to analyze school-level data of eleventh graders in Kentucky high schools and does not find evidence of negative correlations between ACT scores and three- to five-year school overcrowding indicators. **?** is the closest work to mine and analyzes student-level panel data from Wake County. They find that school overcrowding negatively affects students' reading achievement growth, but not students' mathematics achievement growth. They include student, school, and grade-by-year

fixed effects in the regression model to control for time-invariant unobserved characteristics. However, their approach may overlook the effects of time-variant policies, such as annual reassignment plans that switch students to new schools to reduce the overcrowding of existing schools. While these students may experience less severe overcrowding at new schools, their achievement may suffer due to unfamiliar environments and longer commute times, resulting in overcrowding effects in the opposite direction (WRAL, 2018a).

Another implication of overcrowding is larger class sizes or slow implementation of class size reduction mandates. Research on multiple school districts reports that facility constraints due to overcrowding can be a significant obstacle to rolling out class size reduction initiatives (Bohrnstedt and Stecher, 1999; Green and Doran, 2000).[3] There has been extensive literature on the effect of class size on various outcomes, especially for the Tennessee STAR experiment. In general, studies have found strong and positive effects of smaller kindergarten class sizes on student test scores, high school completion, college attendance, wages, fewer arrests, and an array of other adult outcomes such as participation in retirement saving plans (Chetty et al., 2011; Finn and Achilles, 1990; Krueger, 1999; Krueger and Whitmore, 2001). Other studies have found mixed evidence of the effect of smaller class sizes on student outcomes, attributing the reduction in teacher quality to the dampening of small class effect (Hoxby, 2000; Jepsen and Rivkin, 2009).

It is common practice for schools to use temporary or mobile classrooms to address overcrowding (NCES, 2007). Descriptive research has shown this practice to be linked with questionable health and safety conditions and negative student attitudes, despite no significant reported effects in student achievements (Callahan et al., 1999; **?**; NCES, 2007).

The literature has examined two remedies for school overcrowding—mobile classrooms and multi-track year-round calendars. The body of evidence on mobile classrooms mainly comes from qualitative analyses or descriptive statistics. They suggest these classrooms have no effect on student test scores but are associated with negative student attitudes and concerns from teachers and principals on health and safety conditions (**?**).

The effects of multi-track year-round calendars have been studied with more sophisticated quantitative methods. A multi-track year-round calendar allows more students to use facilities by operating on separate schedules with periodic breaks. On the one hand, the alternating schedules help maximize school facilities' utilization and serve more students. On the other hand, they disrupt traditional school schedules and add stress to teachers. Graves (2010) uses panel data methods to analyze school-level longitudinal data from California. She finds a negative impact of multi-track year-round calendars on student academic performance with a 1-2% drop in national percentile rank. While McMullen and Rouse (2012) find no discernible overall impact of year-round calendars on student test scores using fixed-effects models, McMullen et al. (2015) use quantile regressions and show small positive effects for the lowest-performing students.

[3] Table 1.1 on Wake County elementary school characteristics shows that overcrowded schools have slightly higher average class sizes.

1.2.1.2 Estimating School Choice Effects Using Centralized Assignments

A growing literature evaluates education programs by exploiting variation created through randomized lotteries at selective schools or in centralized assignments of a set of schools. Using assignment outcomes based on exogenous variation allows for causal identification of student achievement gains free of bias created by student sorting into schools. Several papers take advantage of random variation based on lotteries administered at the school level by selective schools and open-enrollment programs, using lottery outcomes at each school or program as instruments for student enrollment (**?**Cullen et al., 2006; Engberg et al., 2014).

On the other hand, many studies focus on district-wide centralized assignments comprised of multiple schools where students submit a ranked order of schools to the assignment system. A student's first-choice school is the school that she prefers to any other school. Studies of district-wide school choice plans have frequently used lottery outcomes at students' first-choice schools as instruments for first-choice school attendance. In general, attending the first-choice school in centralized assignments is found to result in positive test score impacts and, in some cases, better outcomes related to crime and college attendance (Hastings et al., 2012; Deming, 2011; Deming et al., 2014). Other studies instrument exposure to charter schools with charter offers, with controls for every combination of charter school lotteries each student was subject to (Abdulkadiroğlu et al., 2011; Angrist et al., 2013, 2016). However, prior research only uses random variation generated at the first-choice schools or oversubscribed schools, limiting the representativeness of the estimates. In centralized assignments, students losing lotteries at oversubscribed schools can create random variations at lower-ranked schools, which are not fully exploited in the studies mentioned above.

Abdulkadiroğlu et al. (2017a) provides a new method that fully exploits the random variation created by lotteries through propensity scores, which represents the probability of receiving an offer from any school on a student's submitted preference list. They suggest that conditioning on propensity scores sufficiently ensures that offers are independent of student characteristics. Using administrative data on Wake County's centralized assignment for magnet school seats, Dur et al. (2020) follow this new approach and causally identify the effect of magnet school attendance for Wake County applicants. They find robust evidence of positive magnet effects on math skills and attendance among magnet applicants. Abdulkadiroğlu et al. (2020) use this method to generate bias-free estimates of school effects on applicants to a minority of New York City high schools that are subject to random assignment. They use these estimates as benchmark results to test the accuracy of two other models developed based on the district-wide sample including non-applicants.

The majority of research on school choice evaluates the programs as a whole and estimates the school choice effects on all students regardless of their fallback options when they lose the lottery. Little research has been done to study policy-relevant aspects of school choice, such as differential effects related to overcrowding. This paper follows the econometric method in Dur et al. (2020) and adapts it to account for differential treatment effects for two student subgroups that vary by their base school overcrowding status.

1.2.2 School Overcrowding in Wake County

The Wake County Public School System is the 15th large school district in the United States, with 159,549 students enrolled in 187 schools as of the 2018-19 school year. Wake County has experienced rapid population growth in recent years. From 2004-2005 to 2015-2016, the student population in the district grew by 38% (NCPS, 2015a). NCPS (2016) concludes that Wake County is in the greatest need of new schools among all school districts in North Carolina. Elementary schools have been identified as the grade levels with the highest need for new construction or existing renovation. In this study, I focus on overcrowding in elementary schools, as it is more prevalent among a larger set of schools.

In 2015, 59 of the 107 (27%) Wake County elementary schools were above 100% of the school's capacity and 29 were above 106%, the threshold by which the National Center for Education Statistics deems a school overcrowded (WCPSS, 2015a; NCES, 2007). An increase in class sizes due to school overcrowding is less of a concern in Wake County, North Carolina. The current North Carolina class size mandate caps are set at 23 students per classroom for grades K-3. With the new state class-size mandate, which limits class size to 19 to 21, going into effect in the fall of 2019, the school district is under even more pressure to increase the number of classrooms. The school district has addressed overcrowding in several ways.

Despite the high costs, Wake County has been building new schools to address the rapid growth in student population. The Community Improvement Program (CIP) in Wake County was passed to allow for $565 million in bonds to be invested in the construction of new schools between 2013 and 2019. (WCPSS, 2019a). Utilizing facilities at new schools to alleviate overcrowding at existing schools often leads to students being reassigned to new schools. Many students and parents subject to reassignment have voiced their concerns, including longer commute times, unexpected changes in the school calendar, or adapting to a new environment(WRAL, 2018a,c,b).

In addition to building new schools, Wake County Public Schools have taken advantage of adapting existing spaces by using mobile/modular classroom units and multi-track year-round schools. As of 2009, there are more than 1,000 mobile units throughout the district. In the 2014-2015 school year, 17.3% of all capacity in elementary school is in temporary classrooms. In 2006, year-round schools were introduced as another remedy for school overcrowding. As of 2016, 37 elementary schools are year-round schools, and seven are multi-track. Multi-track year-round schools split students into four groups with three in class at all times as a way to increase the campus capacity.

Compared to the solutions discussed above, the existing magnet school program helps alleviate overcrowding on a smaller scale, but it has two advantages. First, participation in the magnet program is entirely voluntary—students are free to apply for reassignment at any magnet school or not apply at all. This feature minimizes controversy compared with one-size-fits-all plans to convert school calendars or reassign current students to newly built schools. Second, the assignment operates within the given capacities of each magnet school. Since no additional seats at magnet schools are required, the district can save on new school construction or new mobile classrooms.

1.3 Assignment, Deferred Acceptance, and Assignment Propensity Scores

1.3.1 Centralized Assignment in Wake County

In Wake County, all students are initially assigned to a neighborhood base school based on their home address. The base school is the default enrollment option if students do not apply to be reassigned to a magnet school, an alternative calendar school, or a charter school. Wake County conducts an annual centralized assignment that allows students of any grade to apply to be reassigned to a magnet school. The magnet school program is the most established diversity program in the school district, and is intended to "reduce high concentrations of poverty" and "promote diverse populations" (WCPSS, 2019b). The school district also intends to alleviate school overcrowding with school choice by "maximizing use of school facilities."

Since the 2015-2016 school year, Wake County has used the Deferred Acceptance (DA) algorithm to assign applicants to most district schools. In the assignment process, magnet seats are assigned up to schools' pre-determined capacities based on student preferences and school priorities. Students submit a list of up to five schools ranked by their preference over the schools. Schools also have "preferences" over students, called "priorities", which are functions of student characteristics. In addition to priorities, schools also use a single lottery number assigned to each student to break ties. The lottery numbers are unknown to applicants.

The school district designs the priority structure. At elementary schools, current magnet seat holders in grades K-4 have guaranteed seats in the next academic year. The highest priority at a school is assigned to those with siblings currently attending this school. Students residing in an area with high socioeconomic status receive the second-highest priority. Next, there are a set of facility utilization (or overcrowding) priorities granted to students with overcrowded base schools. The district's intention of addressing overcrowding with school choice is reflected in the utilization priorities. A student with an overcrowded base school has higher priorities at all magnet schools than a student with a non-overcrowded base school, holding sibling status and residence area the same. A student with a severely overcrowded base school is also granted higher priorities than a student whose base school is moderately overcrowded. Overcrowding is defined as having projected enrollment over 105% of projected capacity.[4]

1.3.2 Assignment algorithm: Deferred Acceptance (DA)

Deferred Acceptance (DA) is the most widely-used centralized assignment mechanism. DA is an algorithm with inputs of preferences, priorities, and school capacities coupled with lottery numbers for each applicant, and an output of a vector of assignment outcomes that specifies each student's assigned

[4]There are four categories of overcrowding priorities. The highest priority category is granted to students with base schools projected to be over 120% overcrowded. In 2015-2016, 30 elementary schools were projected to have an overcrowding percentage over 120%. Students with base school overcrowding measure between 115%, and 120% have the second-highest category of utilization priority. The third category is for those from base schools with overcrowding measures between 110% and 115%, and the fourth category for overcrowding measures between 105% and 110%. Many magnet applicants have overcrowding priorities in Wake County. In the 2015-2016 school year's magnet assignment, 44% of applicants have overcrowding priorities, while 62% of those receiving magnet offers have overcrowding priorities.

school (or lack thereof) among the set of schools subject to DA. The DA process is simple and shown as follows:

Step 1. Each applicant applies to the highest-ranked school on her preference list, i.e. the most preferred school.

Step 2. Each school receives the applications and ranks applicants by priorities. Students with the same priorities, if any, are then ranked by lottery numbers. According to the ranked list of all applicants, schools give temporary seats to the highest-ranked up to its capacity and reject others.

Step 3. Each rejected applicant applies to the next highest-ranked school on her list.

Step 4. Each school pools the new applicants together with all students currently offered a temporary seat and ranks them first by priority and then by lottery number. Schools give temporary seats to the highest-ranked applicants up to its capacity and reject others.

Repeat Step 1 through Step 4 until:

Step n. All students are either temporarily seated at a school or have been rejected by all schools on their preference lists. As a result, no one initiates new applications. Now, DA is terminated, and schools offer seats for those who were previously temporarily seated. Each student is seated at only one school since students cannot apply for other schools unless they are rejected at the one they have just applied to.

In Wake County, the school-student-specific priorities are coarse, meaning that many students can share the same priorities at a given school. Therefore, students can be subject to tie-breaking using lottery numbers in order to obtain a magnet seat. When a school is oversubscribed, students with the highest priorities will always be assigned regardless of their lottery numbers. Lottery tie-breaking takes place when a school has more applicants with the same priorities than the number of remaining seats. These applicants are conditionally seated, which means that their assignment is purely determined by lottery numbers. Therefore, conditionally seated applicants at a given school have the same probability of assignment to that school.

1.3.3 Assignment Propensity Scores

It is straightforward that DA allows for random variation for conditionally seated students at oversubscribed schools. Under different draws of random lottery numbers, these students can experience various assignment outcomes. More interestingly, DA also results in random variation at undersubscribed schools.

Imagine a simple assignment with three students and two schools. Students i, j, and k have the same priorities at both schools. All three students prefer school a over school b. School a has a capacity for only one student while school b has more seats available than the number of applicants. In DA, all students apply to the oversubscribed a, and each has a 33% chance of getting in. Two students are rejected and then apply to b, where they are automatically accepted. I can calculate the probability of being assigned to b for all students as $(1-33\%)*100\% = 67\%$.

Thirty-three percent and 67% are examples of assignment propensity scores. A propensity score, or assignment probability, is defined as the probability of being assigned at a school. In the context

of magnet assignments, it is useful to define a sector-specific propensity score over a set of schools $s = 1, ..., S$:

$$PS_i = \sum_{s=1}^{S} Pr(D_i(s) = 1)$$

with $D_i(s)$ indicating the assignment of student i to school s. The magnet propensity score for a student is the sum of her scores for each magnet school.

The propensity score can be calculated based on simulations using different random draws of lottery numbers. The simulated score of an applicant is equal to the proportion of simulations where lottery numbers are drawn in a way that allows the student to be seated at magnet schools. It is worth noting that simulations input preferences, priorities, and capacities in the *original* assignment algorithm that produces offers to students to consider. To generate propensity scores in this paper, I ran the assignment one million times, each time with a new set of random lottery numbers. I then count the number of times students were assigned to each magnet school and divide them by one million to calculate the school-specific propensity scores.

1.4 Empirical Strategy

I follow the econometric method proposed by Abdulkadiroğlu et al. (2017a) and first implemented in the Wake County setting by Dur et al. (2020). This method fully exploits the conditional random variation in assignment outcomes by conditioning on assignment propensity scores, which group students together based on their preferences and priorities. As long as students have the same propensity score, their offer is independent of other characteristics. Compared to lottery indicators used in traditional approaches, such as the first-choice instruments approach, conditioning on propensity scores generates a richer sample, as lottery indicators only group students based on whether they are subject to the same lottery.

Imbens and Angrist (1994) have shown that the local average treatment effect (LATE) for those who comply with their assignment can be causally identified under the assumptions of independence, exclusion, and monotonicity. These assumptions are listed below for a variation of the potential outcomes model. Independence asserts that the assignment outcome must be independent of potential enrollment and outcomes. In the case of centralized school assignments described in Section 3, the assignment outcome is only conditionally random for students who are in the same lottery. Traditional methods only account for random variation at oversubscribed schools that are listed as students' first choice. However, it is embedded in the DA algorithm that undersubscribed schools are also subject to conditional random variation in student offers, since lottery losers apply to a lower-ranked school conditional on losing the tie-breaking with a set probability.

Rosenbaum and Rubin (1983) showed that offers are independent of potential individual characteristics conditional on propensity scores, thus eliminating omitted variable bias. The method in Abdulkadiroğlu et al. (2017a) builds on this. It runs a classic IV model conditional on assignment propensity scores at magnet schools instead of lottery indicators. Propensity-score conditioning al-

lows researchers to identify causal effects based on random variation even at undersubscribed magnet schools. Therefore, this method provides the largest sample size available in the dataset and identifies LATE for a larger and more representative complier sample. Under standard assumptions of LATE, an IV model that follows Abdulkadiroğlu et al. (2017a) is:

$$Y = \mu + \beta D + G(PS) + X'\delta + \epsilon \tag{1.1}$$

with D, magnet attendance, instrumented by Z, an indicator for receiving an offer from any magnet school. $G(PS)$ indicates a function of assignment propensity scores. β is the average effect of magnet attendance on student outcomes for compliers.

In order to study differences in the effect of magnet attendance by base school overcrowding status, I first consider the potential outcomes model for two groups of students:

$$Y = Y_1 + (Y_1 - Y_{0A})(1-D) * A + (Y_1 - Y_{0B})(1-D)] * B \tag{1.2}$$

A is a binary variable that equals one if the student is from an overcrowded base school. $B \equiv 1 - A$ is a binary variable that equals one if the student is from an non-overcrowded base school. Y_{0A} denotes the potential outcome if a student does not attend a magnet school and has an overcrowded base school. Similarly, Y_{0B} represents the potential outcome for a student with a non-overcrowded base school without magnet attendance. Y_1 denotes the potential outcome if this student attends a magnet school. Let D_1 denote a student's potential magnet enrollment with a magnet offer. D_{0j} ($j \in A, B$) indicates her potential enrollment without a magnet offer that is specific to the overcrowding status of her base school..

I impose the standard LATE assumptions—independence, exclusion, and monotonicity. The independence assumption holds when the instrument, the assignment offer, is independent of student characteristics conditional on assignment propensity scores. Random tie-breakers ensure conditional independence in centralized assignments. The exclusion assumption suggests that the instrument cannot directly affect the potential outcome. This assumption holds if magnet offers only affect student outcome by changing student' magnet attendance status for each fallback option group. The monotonicity assumption holds as long as receiving a magnet offer weakly improves the chance of attending a magnet school. It is reasonable to assume that magnet offers do not deter students from attending.

Given that the three LATE assumptions hold, under the potential outcomes model in equation 1.2, the standard IV model presented by equation 1.1 identifies an average of two LATEs for two groups of students by their base school overcrowding status, weighted by their proportion in the complier population, ω_A and ω_B ($\omega_A + \omega_B = 1$):

$$\begin{aligned} \frac{E[Y|Z=1,G(PS)] - E[Y|Z=0,G(PS)]}{E[D|Z=1,G(PS)] - E[D|Z=0,G(PS)]} &= E[Y_{1A} - Y_{0A}|D_{0A} < D_{1A}]\omega_A \\ &\quad + E[Y_{1B} - Y_{0B}|D_{0B} < D_{1B}]\omega_B \end{aligned} \tag{1.3}$$

with

$$\omega_A = \frac{Pr(D_{0a} < D_{1a}|G(PS))}{Pr(D_{0a} < D_{1a}|G(PS)) + Pr(D_{0b} < D_{1b}|G(PS))}, \omega_B = 1 - \omega_A$$

In order to separately identify the two counterfactual-specific LATEs by students' base school overcrowding status, I further impose an additional assumption of constant effects, that is, $Y_{1i} - Y_{0Ai} = \beta_A$ and $Y_{1i} - Y_{0Ai} = \beta_B$ for each observation i. To model two fallback options for magnet applicants, I add an interaction of magnet attendance and an indicator base school overcrowding status in this IV model following the literature (Deming, 2014; Hastings et al., 2012). Then, the model becomes:

$$Y = \mu + \beta_0 D + \beta_A D * A + \gamma A + G(PS) + X'\delta + \epsilon \tag{1.4}$$

with D and $D*A$ instrumented by Z and $Z*A$. A denotes a student's base school overcrowding status. In this model, β measures the mean magnet effect for complying students from a non-overcrowded base school, and β_A measures the differential gain from magnet attendance by base school overcrowding status.[5] While β_A cannot be interpreted as the true overcrowding effect, it provides a novel alternative measure of overcrowding effects on student achievements by evaluating how students from overcrowded schools differentially respond to the magnet school program.

1.5 Wake County Data and Descriptive Statistics

This study proposes to estimate the causal impact of school overcrowding through student-level administrative data merged with student assignment outcomes and school-level indicators on overcrowding. I study school overcrowding effects in elementary schools, which has been identified by Public Schools of North Carolina as the school type most in need of new and renovated facilities (NCPS, 2016).

I use assignment and enrollment data on grade-entry students in the 2015-2016 school year, who submitted their magnet applications in February 2015. These students are followed through the end of their second grade in the 2017-2018 school year, where student outcome data are recorded. I focus on students who were kindergarteners in the 2015-2016 school year in order to allow the sample to contain the largest number of students in one grade. Twenty-two percent of applicants for magnet elementary schools were entering kindergarten, the most among all elementary grade levels. Restricting the sample to grade-entry magnet applicants likely results in estimating magnet effects based on a more representative sample of students. However, grade-entry students may have different characteristics and gains from magnet attendance compared to non-grade-entry students. Students in the latter group likely have prior experience at traditional schools or have recently moved into the Wake County school district.

[5]Table A.1 compares differences in student characteristics by magnet offers and base school overcrowding status with controls for assignment propensity scores. Propensity scores seem to provide sufficient control to allow for relatively balanced student characteristics for lottery losers and winners. Lottery losers with a non-overcrowded base school are more likely to be African American. Lottery winners with an overcrowded base are less likely to have siblings or having an older sibling that also attends Wake County public schools.

1.5.1 School Characteristics

The school-level administrative data has information on actual overcrowding levels in the 2014-2015 school year and projected overcrowding levels in the 2015-2016 school year. The overcrowding measure is calculated by dividing the number enrolled students by the school's capacity. The planned facility utilization capacity is based on the number of permanent classrooms and the "optimal" number of temporary (mobile) classrooms.[6] The actual facility utilization capacity adjusts the planned capacity by accounting for the actual number of temporary classrooms used and other school-specific changes in its needs for classrooms. In this paper, I focus on the actual overcrowding level in the 2014-2015 school year, which is calculated by dividing each school's 2014-2015 enrollment by its actual capacity for the school year.[7] I construct a binary indicator for school overcrowding that equals one if a student's base school is over 105% overcrowded. To leave out the effect of mobile classrooms, my calculation is based on the capacity of permanent classrooms only by removing temporary classroom capacity from the total school capacity.[8] Throughout this paper, I consider all magnet schools as non-overcrowded schools, given the fact that the school district wishes to move students from overcrowded traditional schools to magnet schools.

The school-level data also contains information on school demographics and academic performance. Table 1.1 shows differences in school characteristics by overcrowding status. Magnet schools are listed separately in Column 3. Overcrowded schools are generally larger schools and have approximately one additional student in an average classroom. However, overcrowded schools also have slightly higher teacher-to-classroom ratios. It is important to note that overcrowded schools have higher teacher quality in terms of experience than non-overcrowded schools and magnet schools. Overcrowded schools also have substantially lower numbers of acts of violence and short-term suspensions in addition to slightly higher attendance percentages. Also, overcrowded schools score significantly higher on average in the school achievement scores calculated based on the average student EOG math and reading test scores in the 2014-2015 school year. This finding further motivates this paper's empirical methodology that allows Y_0 to differ based on base school overcrowding status. However, overcrowded schools have higher ratios of students to internet-connected computers, which may indicate old or deteriorating equipment.

1.5.2 Student Characteristics

The student-level assignment data has detailed information on all magnet applications, including student preferences, priorities at each school, lottery numbers, school capacities, and assignment outcomes. To calculate the propensity score for each student, I construct a new dataset using preferences, priorities, and capacities in the actual assignment and a set of lottery numbers that are randomly re-

[6] The optimum number of temporary units is defined as the number of mobile classrooms that meets a number of criteria, including being supported by dining and toilet facilities and specialized educational program spaces such as gym.

[7] Projected overcrowding is calculated by dividing the projected enrollment in the next school year by the planned capacity. Projected overcrowding serves as the ground for assigning utilization priorities in the Wake County magnet assignment.

[8] Another overcrowding indicator was created for the case where the school is over 105% overcrowded even with temporary classrooms. I also considered projected overcrowding, overcrowding measures on high class size and temporary classrooms usage. Results using these overcrowding measures are reported in Table A.4.

drawn from a uniform distribution between zero and one. I then use this dataset as the input to run the original assignment algorithm and obtain a new set of assignment outcomes for each student. Due to lottery numbers changes, a student may switch between assigned and unassigned status or receive a different magnet assignment in the simulation. I then repeat this process one million times using the High Performance Computing cluster, thus collecting 1 million assignment outcomes for each student. To aggregate simulation results, I calculate each student's "simulated propensity score" for magnet assignment by counting the number of simulations where they are assigned to any magnet school and dividing it by 1 million.[9] Students with simulated propensity scores strictly between 0 and 1 are subject to random variation in assignment outcomes.

The student-level administrative records also include detailed information on student demographics such as race and gender, enrollment, and non-cognitive outcomes such as absences. Based on students' home addresses, I create an indicator for whether students live in residence with other students attending a Wake County magnet school ("siblings"). By comparing grades of students living in the same address, I also record whether a student has older siblings attending a magnet school. As for cognitive outcomes, I focus on two test scores that measure early literacy and math skills in the data.[10]

The Number Knowledge Test (NKT) and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are individually administered to grade K-2 students three times in the beginning, middle, and end of each academic year.[11]

The Number Knowledge Test (NKT) is an oral test designed to measure intuitive knowledge of numbers and basic arithmetic skills such as counting, adding, and subtracting that an average student has at each age level (Case et al., 1996). The test contains thirty questions in four levels corresponding to the knowledge acquired by a typical student at the age of 4, 6, 8, and 10, respectively. The student cannot proceed to the next level of questions if she has committed three consecutive errors in the current level. The NKT score is widely used to measure early math achievement in literature from various disciplines (Jerrim and Vignoles, 2016; Ribner et al., 2019; Zhang et al., 2019). Several studies have shown a strong positive correlation between the NKT score and later math and reading attainment and a negative correlation with behavioral problems such as physical aggression in elementary school (Cowan et al., 2011; **?**; Jordan et al., 2007; Romano et al., 2010). Gersten et al. (2005) finds that the NKT works well as a screening measure for students with math difficulties. Research using growth models shows students in low-increasing trajectories of early number knowledge consistently fall behind in math learning through elementary school, suggesting the importance of early identification and intervention for students with

[9] Unfortunately, the analytic DA propensity score proposed by Abdulkadiroğlu et al. (2017a) is not applicable in the setting of the Wake County assignment. In the 2015-2016 school year magnet assignment, Wake County performs an assignment rule that only fills 90% vacant seats after assigning guaranteed seats to students with older sibling priorities. The remaining 10% of seats are assigned based on a lottery without accounting for student priorities. As a result, the Wake County assignment rule results in two lotteries at each school if capacities are not filled after assigning guaranteed seats to students with older sibling priorities. The analytic DA score is no longer applicable since it assumes that only one lottery is carried at each school to assign students.

[10] This is because the grade-entry cohort in elementary schools have not yet taken the end-of-grade standardized tests, which begins in the third grade.

[11] Since grade-entry students have no recorded cognitive outcomes before Kindergarten, I consider test scores at the beginning of Kindergarten as baseline test scores.

deficits in number sense following NKT test results (**?**Jordan et al., 2007). The raw test score is between 1 and 30. I create an indicator for students who score above 14, the cutoff for having math skills, not below the grade level of K-1.[12] This indicator of not scoring below grade level is interpreted as the student not being at-risk in terms of early math skills, as low NKT scores are predictive of low achievements in math learning as students progress to higher grades (**?**Jordan et al., 2007).

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is designed to measure early literacy skills and identify students experiencing difficulty acquiring these skills to provide support early on (Good and Kaminski, 1996). Like the NKT, the DIBELS is also an individually administered oral test where students read out passages to teachers. For second graders, DIBELS consists of measures on students' speed and accuracy in reading connected text (Oral Reading Fluency, ORF) and comprehension of the text (Retell). DIBELS has gained widespread use in the United States, with numerous studies showing its predictive power for later reading attainment in elementary school and its effectiveness with identifying students at risk of reading difficulties (Good et al., 2001; Smolkowski and Cummings, 2016; Pears et al., 2016; Roehrig et al., 2008). More specifically, Barger (2003) find a strong correlation between the DIBELS ORF score and North Carolina end-of-grade reading scores when these two tests were taken one week apart. I create a binary indicator for not scoring well below the DIBELS test benchmark and focus on this outcome for literacy skills in this paper. This definition is consistent with the test's objective to identify students with literacy skills "at-risk" of falling behind.

A large body of literature has examined non-cognitive skills that are not captured by standardized tests and shown that these skills have significant effects on long-term outcomes in students' adult lives (Heckman and Rubinstein, 2001; Borghans et al., 2008; Lindqvist and Vestman, 2011). Following the literature, I use one behavioral measure available in the dataset as a proxy for non-cognitive skills - the log of the number of absences between grade 0.[13]

Table 1.2 shows student characteristics for all students who applied for a Kindergarten seat in the magnet school program for the 2015-2016 school year. Column (2) and (3) show mean characteristics for subgroups by base school overcrowding status. Overcrowded schools have a slightly higher proportion of male students. Students from overcrowded schools are less likely to be African American and and more likely to be Hispanic students compared to those from non-overcrowded schools. Average NKT scores at the beginning of Kindergarten is slightly higher at overcrowded schools, while the proportion of students not at-risk in literacy skills is lower. I consider test scores recorded at the beginning of Kindergarten as baseline test scores that indicate student ability before any school influence. Hence, these results suggest that both positive and negative sorting into overcrowded schools depending on the subject. As for assignment outcomes, students from overcrowded base schools are more likely to receive an offer and have higher median priority at first choice school compared to students from non-overcrowded base schools. This result is consistent with the priority structure design that prioritizes the magnet

[12]The developmental conversion chart that exhibits how raw scores are converted to grade levels is available at https://s3.amazonaws.com/ecommerce-prod.mheducation.com/unitas/school/explore/sites/number-worlds/number-worlds-number-knowledge-test.pdf.

[13]In order to measure non-cognitive skills, Jackson (2018) constructs a behavior index with five behavioral variables, including log absences.

assignment of these students.[14]

1.6 Results and Discussion

1.6.1 Overall effects

The estimation sample for Table 1.3 is all students with an assignment propensity score strictly between 0 and 1 who do not have a magnet as base school and are continuously enrolled between 2015-2016 and 2017-2018. School-level controls for base schools include binary indicators for below-median one-year teacher turnover rate, as well as average acts of violence and short-term suspensions. These measures are recorded in the 2014-2015 school year, prior to the enrollment of Kindergarten magnet applicants for the 2015-2016 school year.[15]

I measure math skills for second graders using scores from a shortened version of the Number Knowledge Test (NKT). The mean score is 25.2 for second graders at the end of the 2017-2018 school year in the sample. This score's chronological age equivalent is about seven to eight years old, and the grade level equivalent is about grade 2-3 (Case et al., 1996). Of the 973 children in the estimation sample, only 9 had missing NKT scores. I also create an indicator that equals one if a student has an NKT score above 14, suggesting a grade level equivalence of grade 1-2 and above.

I present in Part A, Table 1.3, the second-stage estimates of magnet effects on math skills by school overcrowding status based on the model shown in equation 2.4. Column (1) shows the differential magnet effect, which I consider as an alternative measure of overcrowding effects. If falling back to an overcrowded base school instead of a non-overcrowded base school results in worse outcomes, i.e. negative overcrowding effects, the magnet gain for students from overcrowded schools should be stronger. Column (2) and (3) show that magnet effects are positive and statistically significant regardless of applicants' base school overcrowding status. Results suggest that attending a magnet school increases NKT score by more than four points on average for applicants regardless of their base schools' overcrowding status. These points gain are equivalent to an improvement in math skills equivalent to approximately one grade level. Magnet effects are also positive and statistically significant on not receiving the at-risk status of math skills. Magnet attendance contributes to a 31 to 33 percent reduction in at-risk status in math skills, depending on base school's overcrowding status. The differential magnet gain for both math

[14] Table A.2 shows that lottery winners with an overcrowded base school are slightly more likely to attend an offered magnet school than lottery winners with a non-overcrowded base school. Lottery losers with an overcrowded base have a higher rate of compliance by attending their base school than lottery losers with a non-overcrowded base school. In general, students with an overcrowded base are also less likely to leave the school district. Figure A.1 shows the relationship between three student outcomes and magnet attendance by base school's actual overcrowding intensity in the 2014-2015 school year. The outcome difference by magnet attendance for log absences increases as overcrowding measure reaches above 100%.

[15] Table A.2 shows the enrollment status of magnet school applicants who are subject to lotteries. I have left out applicants with a magnet base school. Results are presented for four subgroups by whether they have won the lottery and their base school overcrowding status. A substantial proportion of lottery losers with non-overcrowded schools end up attending a magnet school, compared to the other groups. In addition, applicants with a non-overcrowded base school are more likely to leave the school district regardless of whether they received a magnet seat. Column (1), Table A.3, shows the first-stage estimate of the effect of receiving a magnet offer on magnet school attendance in Kindergarten. Column (2) suggests some level of differential compliance by base school overcrowding status, as evidenced in Table A.2.

outcomes are small and not significantly different from zero, which indicates minimal overcrowding effects on students' math skills.[16]

Part B, Table 1.3, displays magnet effects on normalized Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test scores and being identified as at-risk in the DIBELS test, which measures early literacy skills. Compared to the NKT test, there are around 30 more students with missing scores.[17] For students from both overcrowded and non-overcrowded base schools, attending a magnet school has no discernible effect on improving DIBELS scores or the chance of not being an at-risk student. These null results of magnet effects on literacy skills are consistent with the literature (**?**Dur et al., 2020). The gap in magnet effects by overcrowding status suggests overcrowding effects is not statistically significantly different from zero. These results, however, are also estimated with large noise.[18]

Part C, Table 1.3, shows results on the non-cognitive outcome of absences. The dependent variable, log absences, is calculated by taking the logarithm of one plus all absences of each student in the three school years between 2015-2016 and 2017-2018. Estimates in Column (2) show that students with an overcrowded base schools have a statistically significant reduction in the number of absences by more than 90% from attending a magnet school. Column (3) shows magnet effects of a 40% reduction in absences for students from non-overcrowded base schools. The differential gain is large and statistically significant, indicating strong negative overcrowding effects on student attendance. The effect size is an almost 50% increase in absences throughout the three school years observed in my sample, which is equivalent to attending around 10 days fewer than students in non-overcrowded base schools. Possible explanations for negative overcrowding effects on non-cognitive outcomes include reduced engagement with the school.[19]

1.6.2 Mechanisms

Table 1.4 further explores the path to overcrowding effects by examining two common arguments for negative overcrowding effects, higher class size and the use of temporary (mobile) classrooms. Model specifications are similar to those presented in Table 1.3. I interact measures of class size or mobile classrooms with magnet attendance for students from overcrowded base schools as well as its instrument, receiving a magnet offer for students from overcrowded base schools. Columns (1) and (2) present two overcrowding effects with respect to the overcrowded school's class size. I measure class size using each school's K-2 average class size in the 2014-2015 school year and create a binary indicator

[16] Table A.4 and A.5 show regression results under six alternative specifications that are largely consistent with results shown in Part A, 1.3. For specification checks, I switched to alternative measures of overcrowding status and changed the function form of propensity scores as well as the estimation sample. In general, the differential gains from magnet attendance by overcrowding status are reported as close to zero except for Column (3), Table A.4.

[17] To obtain the most information from the WCPSS dataset, I use mid-of-year DIBELS test results in second grade.

[18] Table A.4 and A.5 presents regression results under five alternative specifications. Overcrowding effects can be positive or negative depending on the definition of overcrowding. Column (4), Table A.4 suggests that the large average K-2 class size of a school is associated with negative effects on literacy skills.

[19] Table A.4 and A.5 show five alternative specifications for non-cognitive outcomes. The negative overcrowding effects on attendance are robust to various functional forms of propensity scores, but not to different definitions of overcrowding measures. Overcrowding effects are not observed for a few overcrowding measures including projected overcrowding, mobile classroom usage, or high class size.

for schools with average class size above the district average of 21 students per class.[20] 14 out of 26 overcrowded schools are defined to have higher-than-average class sizes. Similar to Table 1.3, I do not observe any statistically significant overcrowding effects on math and literacy skills. The strong negative overcrowding effect on student attendance is only observed for overcrowded schools with above-average class size. Similarly, I estimate the 2SLS model using interactions with mobile classroom usage and present two overcrowding effects in Columns (3) and (4). I refer to the facility utilization report for the number of temporary classrooms used in each school in the 2014-2015 school year. Mobile classroom percentage is calculated by dividing the number of mobile classrooms by the sum of permanent and mobile classrooms. The median is 0.11, suggesting 11% of all classrooms in a school is a temporary unit. 5 out of 26 overcrowded schools did not employ mobile classrooms in this school year. Again, no overcrowding effects on math and literacy outcomes are observed. Despite the lack of power, the overcrowding effects on absences is larger for schools that uses mobile classrooms, especially those with an above-median mobile classroom percentage. These results suggest both school characteristics contribute the negative overcrowding effects on student attendance. However, I cannot disentangle the two effects due to lack of variation in mobile classroom usage - all but five overcrowded schools use mobile classrooms to accommodate students. It is also unknown whether the cohort observed were placed in any mobile classrooms between 2016 and 2018.

1.6.3 School Performance and Overcrowding Effects

We may suspect that overcrowded schools have more enrolled students over their capacities because they have specific characteristics that are correlated with student achievement. Table 1.1 suggests that overcrowded schools have higher student performance levels than other schools in Wake County. To address this achievement gap by overcrowding status, I separate the sample into two subsamples by a binary indicator for high academic performance at the student's base school in the school year prior to her enrollment. The indicator equals one if the base school has school achievement scores of A or B for End-of-Grade (EOG) math and reading tests, as indicated by the school report card in the 2014-2015 school year.[21] To receive school achievement scores of A or B, the school must have at least 70% of students being grade-level proficient in the subject (NCPS, 2014, 2015b). School achievement scores are reported on North Carolina School Report Cards as core measures of each school's academic performance. The subsample analysis allows me to identify two sets of differential magnet effects for students based on base school overcrowding. This model also provides additional information on heterogeneity in achievement gains within students with an overcrowded base school.

I separately identify the model shown in Table 1.5 for two subgroups by base school achievement scores in the 2014-2015 school year. Descriptive statistics in Table 1.1 suggest that overcrowded schools, in general, have higher SPG scores, which motivates this subsample analysis to compare magnet effects

[20] The state's class size mandate is 23 between 2014 and 2018.

[21] The EOG math and reading tests are statewide assessments on student performances concerning the North Carolina Standard Course of Study (NCSCS). Elementary school students in grades 3-5 enrolled in public and charter schools participate in EOG tests administered by the North Carolina Department of Public Instruction's Accountability Services Division. An 80% weight is assigned to school achievement scores when calculating the overall School Performance Grade (NCPS, 2014).

by overcrowding status based on base school quality. Here, a high-performance school is defined as having received an "A" or "B" for both EOG math and EOG reading on the school report card in the 2014-2015 school year. In 2014-2015, 32 out of the 107 schools in Wake County have received an "A" or "B" for both subjects on the school report card, including 23 overcrowded traditional schools, three magnet schools, and six non-overcrowded traditional schools. The sample size for students with a high-performance base school is 331.

Table 1.5 shows heterogeneous overcrowding effects by base school's academic performance. Negative overcrowding effects on math skills are observed among students with a low-performance base school. However, these effects are measured with large noise. Overcrowding effects on math are positive among students with high-performance base schools, suggesting high quality of the set of eight overcrowded school among 30 high-performance schools. Both subsamples generate null overcrowding effects on literacy. Negative overcrowding effects on attendance are observed for both subsamples. Consistent with Part A, the effect size for students with a high-performance base school is smaller.

Tables 1.3 to 1.5 show the important aspect of overcrowding effects on non-cognitive outcomes, which is overlooked in the literature. The results also suggest a high degree of heterogeneity in overcrowding effects by base school performance. If policymakers are concerned about maximizing the level of magnet effect, results in this paper imply that they may consider granting higher priorities to students from an overcrowded school with low school achievement scores when designing the priority structure of the magnet reassignment.

1.7 Conclusion

In this study, I take advantage of the centralized magnet school assignment in Wake County, North Carolina, to examine the school choice effects on students who attempt to leave overcrowded magnet schools. I exploit the randomness created by assignment tie-breakers to disentangle school crowding effects for assignment lottery losers. The results suggest mixed effects of school overcrowding on student cognitive and non-cognitive outcomes. Having an overcrowded base school is not associated with stronger magnet effects on math skills measured by the Number Knowledge Test (NKT) and literary skills measured by the Dynamic Indicators of Basic Early Literacy Skills test (DIBELS). Since students from overcrowded base schools do not experience higher gains from "escaping" overcrowded traditional schools than others, I find no evidence of negative overcrowding effects on early cognitive outcomes for magnet applicants in Wake County.

Overcrowding effects on log absences are strong and negative. Students from overcrowded base schools experience significantly higher magnet gains in reducing absences, suggesting a substantial negative overcrowding effect on student attendance. I have examined two potential mechanism for this overcrowding effect. Although the second graders in my sample were subject to a statewide class size mandate of 23 students per classroom at the time, larger class sizes are still associated with large overcrowding effects on attendance. The use of temporary classrooms in overcrowded schools is also linked with strong overcrowding effects. The subgroup analysis of overcrowding effect by base school's

academic performance indicates heterogeneity in both the quality of overcrowded schools and overcrowding effects.

Many studies have suggested a negative correlation between early absenteeism and academic achievement trajectory, including test scores and high school dropouts (Alexander et al., 1997; Brocato, 1989; Krohn and O'Connor, 2005; Lin and Chen, 2006). Overcrowding in elementary school can have long-term impacts on both cognitive and non-cognitive outcomes. Utilizing an existing magnet program to address overcrowding may be part of the solution.

Table 1.1 Elementary School Characteristics by School Overcrowding Status

School characteristics	Overcrowded traditional schools	Non-overcrowded traditional schools	Magnet schools
	(1)	(2)	(3)
Number of students	837.462	684.500	565.360
	(198.350)	(194.577)	(120.878)
Average grade 0-2 class size	21.218	20.327	20.393
	(1.107)	(1.503)	(1.798)
Number of teachers per classroom	1.223	1.112	1.122
	(0.151)	(0.204)	(0.159)
Percentage of teachers with 11+ years experience	0.537	0.491	0.454
	(0.097)	(0.112)	(0.105)
Percent of teachers with masters or higher degree	0.365	0.362	0.368
	(0.075)	(0.074)	(0.056)
High math achievement scores	0.769	0.463	0.360
	(0.430)	(0.503)	(0.490)
High reading achievement scores	0.654	0.389	0.200
	(0.485)	(0.492)	(0.408)
Average daily attendance percentage	0.962	0.958	0.959
	(0.004)	(0.005)	(0.004)
Number of crimes or acts of violence per 100 students	0.052	0.146	0.146
	(0.074)	(0.184)	(0.230)
Short term suspensions per 100 students	2.111	3.392	5.766
	(4.097)	(3.290)	(7.853)
Observations	26	54	25

Note: Author's calculations based on the 2014-2015 School Report Card provided by the North Carolina Department of Public Instruction and the 2014-2015 Facility Utilization Report provided by the Wake County Public School System (WCPSS, 2015b). Information in School Report Cards include school-level data student performance and academic growth, school and student characteristics. High math or reading achievement scores refers to the school scoring an "A" or "B" on the School Report Card in the given subject.

Table 1.2 Applicant Characteristics by Base School Overcrowding Status

Applicant characteristics	Base school type: Overcrowded traditional school	Base school type: Non-overcrowded traditional school	Base school type: Magnet school
	(1)	(2)	(3)
Received a magnet offer	0.660	0.424	0.475
	(0.475)	(0.495)	(0.500)
Median priority at 1st choice	23	8	16
Male	0.544	0.468	0.521
	(0.499)	(0.499)	(0.501)
African American	0.166	0.233	0.164
	(0.373)	(0.423)	(0.371)
Hispanic	0.127	0.088	0.063
	(0.334)	(0.283)	(0.244)
Other race	0.205	0.186	0.097
	(0.404)	(0.390)	(0.296)
Limited English Proficiency	0.046	0.047	0.021
	(0.211)	(0.212)	(0.144)
Has sibling	0.359	0.327	0.265
	(0.481)	(0.469)	(0.442)
Has older sibling(s) enrolled in the district	0.263	0.273	0.231
	(0.441)	(0.446)	(0.422)
Baseline NKT score	11.485	10.967	11.385
	(3.923)	(3.928)	(3.498)
Baseline DIBELS score not at-risk	0.872	0.901	0.887
	(0.335)	(0.298)	(0.318)
Observations	259	853	238

Note: 26 magnet school applicants have base schools that open in the 2015-2016 school year, Abbotts Creek Elementary School and Scotts Ridge Elementary School. They are removed from the sample.

Table 1.3 Estimates of Overcrowding Effects

Student Outcomes	Difference (Overcrowding effect) (1)=(3)-(2)	Overcrowded base (2)	Non-overcrowded base (3)
Part A: Math			
NKT Score	-0.184	4.796**	4.611***
	(1.619)	(2.186)	(1.294)
NKT Not At-risk	-0.020	0.328***	0.308***
	(0.091)	(0.117)	(0.068)
N		248	716
Part B: Literacy			
DIBELS Score	-0.107	0.224	0.117
	(0.308)	(0.484)	(0.307)
DIBELS Not At-risk	-0.071	0.026	-0.045
	(0.089)	(0.140)	(0.095)
N		243	692
Part C: Non-cognitive			
Log Absences	0.489*	-0.925**	-0.436*
	(0.281)	(0.401)	(0.235)
N		251	722

Note: Sample includes magnet school applicants who are recorded in the Wake County public school data continuously between the 2015-2016 school year and the 2018-2019 school year. Specifications include gender, race, siblings enrolled in the district, four base school characteristics, and nine propensity score controls indicating propensity score rounded to 0.1. Robust standard errors are reported. * p<.10, ** p<.05, *** p<.01.

Table 1.4 Overcrowding Effects by Class Size and Mobile Classroom Usage

	Class size		Mobile Classrooms		
Student Outcomes	Class size above district average	Class size below district average	Mobile classrooms percentage above median	Mobile classrooms percentage below median	No mobile classrooms
	(1)	(2)	(3)	(4)	(5)
Part A: Math					
NKT Score	-0.717	1.111	-1.947	0.541	-0.460
	(1.960)	(2.247)	(2.808)	(2.491)	(0.869)
NKT Not At-risk	-0.075	0.127	-0.262	0.122	0.013
	(0.110)	(0.126)	(0.177)	(0.134)	(0.049)
Part B: Literacy					
DIBELS Score	-0.210	0.232	0.284	-0.751	-0.050
	(0.374)	(0.408)	(0.560)	(0.458)	(0.137)
DIBELS Not At-risk	-0.068	-0.078	0.079	-0.224	-0.050
	(0.108)	(0.128)	(0.160)	(0.149)	(0.039)
Part C: Non-cognitive					
Log Absences	0.665*	0.025	0.771	0.539	-0.079
	(0.355)	(0.362)	(0.534)	(0.417)	(0.111)

Note: Results show the differential magnet gains by base school overcrowding status interacted with base school class size or mobile classroom measures. Sample includes magnet school applicants who are recorded in the Wake County public school data continuously between the 2015-2016 school year and the 2018-2019 school year. Specifications include gender, race, siblings enrolled in the district, four base school characteristics, and nine propensity score controls indicating propensity score rounded to 0.1. Robust standard errors are reported. * p<.10, ** p<.05, *** p<.01.

Table 1.5 Subgroup Overcrowding Effects by School Performance

Student Outcomes	Low-performance School (1)	High-performance School (2)
Part A: Math		
NKT Score	-3.990	2.567*
	(3.460)	(1.468)
NKT Not At-risk	-0.286	0.218**
	(0.179)	(0.093)
N	636	328
Part B: Literacy		
DIBELS Score	0.006	0.065
	(0.707)	(0.237)
DIBELS Not At-risk	-0.069	0.007
	(0.228)	(0.029)
N	610	325
Part C: Non-cognitive		
Log Absences	1.297*	0.227
	(0.694)	(0.250)
N	642	331

Note: Results are estimated for two subsamples based on base school's achievement score in the previous school year. High-performance school is defined as scoring an A or B for math and reading in school performance grade in 2015. Subsamples includes magnet school applicants who are recorded in the Wake County public school data continuously between the 2015-2016 school year and the 2018-2019 school year. The Specifications include gender, race, siblings enrolled in the district, four base school characteristics, and nine propensity score controls indicating propensity score rounded to 0.1. Robust standard errors are reported. * p<.10, ** p<.05, *** p<.01.

CHAPTER

2

ESTIMATING SCHOOL CHOICE TREATMENT EFFECTS: AN ANALYSIS OF ASSIGNMENT PROPENSITY SCORES

2.1 Introduction

Many large urban districts now conduct centralized school assignments. Lotteries are commonly used in assignment mechanisms like deferred acceptance and Boston mechanism to ensure fairness in allocating seats at oversubscribed schools to applicants. Therefore, students with the same preferences and priorities at a given school have the same probability of receiving an offer for a seat. The use of lotteries also results in the conditional independence of assignment outcomes, creating random variation helpful for identification of school choice effects.

Abdulkadiroğlu et al. (2017a) proposed a new method to fully exploit the quasi-experimental variation using conditional probability of random assignment, or propensity scores, as controls in student outcome models. If a student has a propensity score strictly between zero and one, then she contributes some random variation that can be used for identification. Apart from determining the maximal number of students that enter into the model estimating school choice treatment effect, propensity scores also reflect the relationship between preferences and priorities in an assignment.

While there is a growing literature on the evaluation of school choice treatment effects, this is the first paper that discusses details on how propensity scores are generated and the implications of the distribution of different propensity scores on identification. The most straight-forward impact of a

propensity score distribution is on the sample size because only those with a propensity score strictly between 0 and 1 can be identified. This paper explores various characteristics of an assignment that shapes the propensity score distribution of all students. Then, I discuss the ideal properties of the propensity score distribution for more power and better precision in estimating school choice treatment effects, thus creating a link between real life school choice scenarios and identification. Another purpose of this paper is to provide some intuition on assignment characteristics that are related to "uncommon" propensity scores such as 0.0001 and 0.9999 as well as the small size difference between propensity scores such as 0.01 and 0.001.

2.2 Characteristics of Assignments

A number of characteristics of an assignment can affect the propensity score distribution. Preferences and priority structures are obvious candidates as propensity scores are assignment probabilities conditional on these two features of students. Here, the discussion is limited to the assignment mechanism deferred acceptance (DA). S denotes the set of schools for which propensity score is calculated.

2.2.1 Commonality of Preference

Commonality of preference affects propensity scores because more students compete for seats at the same school when their rankings over schools are more correlated with each other. In real life, it is likely that some schools are known for their high quality and thus become over-demanded in assignments.

2.2.1.1 Common preferences and no priorities

In the extreme case of completely common preference, where every student has the same preference on schools, all students without a seat will apply for the same school in each round of assignment, according to DA. Assume that S only consists of the top school by student rankings in the common preference. If there are no priorities, then the one school in S is heavily over-demanded and the realized assignment completely relies on the set of tie-breakers in the lottery. By the law of large numbers, every student has the same propensity score, which equals an average assignment probability of the fraction of available seats per applicant. As we include more schools down the rankings list to S, the propensity score increases to the fraction of total number of available seats in S per applicant. On the other hand, when preferences are completely random, competition at each school is less fierce and students are highly likely to be assigned to their top choices. For students who rank schools S high, their propensity scores are equal to or close to one. For students who rank schools S low, their propensity scores are equal to or close to zero. Therefore, the propensity score distribution is U-shaped.

2.2.1.2 Common preferences with priorities

Without priorities, students keep applying for schools down their list if schools are oversubscribed. This process stops when they win a lottery and thus obtain a seat. When priorities are introduced,

students with no or low level of priorities can instantly lose to students with priorities when competing for highly-sought-after seats. The introduction of priorities lowers the importance of lotteries when there is competition. If priorities are randomly assigned and not correlated with preferences, then students will be more likely to have a zero or one propensity score, compared to when there are no priorities. Fewer students will have a propensity score that is strictly between zero and one.

2.2.2 Rules of Priorities (assuming identical rules for all schools)

Rules of priorities can affect the number of students that are conditionally seated. As opposed to lotteries, which ensure that similar students are treated equally, priorities serve the purpose of distinguishing students satisfying certain requirements from those who don't. In real life, priorities are often designed by school districts to pursue goals such as diversity.

2.2.2.1 Number of (even) priority classes

For simplification purposes, I consider the case where all priority classes at a given school contain the same number of students, i.e. priority classes are even. For example, if there are 100 students applying and the number of even priority class is 2, then the number of students in each even priority class is $100/2 = 50$. Higher number of priority classes leads to a decrease in the number of conditionally seated students because there can only be one priority class where students are conditionally seated and winners are determined by lotteries. As the number of classes goes up, the propensity score distribution becomes more U-shaped because there are fewer students with p-score strictly between 0 and 1.

2.2.2.2 Number of students in each priority class

The implications of uneven priority structures depends on the proportion of students in each priority class and the proportion of S's capacity over all schools' combined capacity. Consider the simplest case where preferences are common. When all students apply for the same school, the conditionally seated priority class can be derived through comparing the sum of proportion of students in top priority class and the proportion of S's capacity. Suppose that one priority structure dictates that 60% of students have no priorities, 30% of students have low level priorities, and 10% of students have low level priorities (denoted by [0.6;0.3;0.1]). If the capacity of S takes up 50% of all capacities, then the conditionally seated priority class is the class with no priorities because $50\% - (30\% + 10\%) > 0$.

2.2.3 Correlation between preferences and priorities

When preferences and priorities are highly correlated, students have high priorities at their favorite schools. For students who prefer schools in S, their chance of being assigned is higher than that when priorities are random. Therefore, there are more propensity scores that are close to or equal to one. For students who prefer schools that are not in S, it is more likely that they will be assigned to these schools rather than schools in S, resulting in propensity scores that are close to or equal to zero. Compared to random priorities, priorities that are correlated with preferences leads to more U-shaped distributions.

2.3 Assignment models on simulations

The propensity score for DA can be estimated by computer simulations. By repeatedly drawing lottery numbers and conducting assignments, assignment outcomes are in effect drawn from the relevant permutation distribution. The propensity score is computed through taking the average of assignment outcome across draws.

In order to evaluate the origins of differences in propensity score distributions, I set up a model for assignments that easily allows for various assignment features. The purpose of this step is to create an environment with similar features to real-life school assignments. In doing so, I manually control for and change parameters of the assignment and explore their effects on the resulting propensity score distribution.

To utilize and evaluate the simulated propensity scores for identification, I also model for student responses after receiving an assignment outcome. Student responses include students' choice to attend magnet schools after receiving an offer, as well as students' academic outcomes measured by test scores. Both are crucial in treatment effect identifications of public school programs. Modeling student attendance and performance allows me to evaluate treatment effects using propensity scores and comparing estimation results with the true value. Combined with the assignment models, this model informs researchers about the identification implications of certain assignment rules carried out by school districts.

There are three stages in my model for magnet assignment, namely assignment, enrollment, and achievement. At stage 1, students indexed by $i = 1, ..., N$ decide whether to apply to each of S magnet schools, indexed by $s = 1, ..., S$. One important determinant of student preference is perceived potential score gain of schools. Students compare the school in question with their outside option and decide on applying. The binary variable A_{is} indicates that student i applies to school s. Let A denote the set of information that is preference over schools of all applicants. School s has priorities P_{is} for student i. Let P denote the set of information that is priorities of all schools over all students. For simplicity, I assume that $q = N/S$ is the number of seats in each school and that all schools are acceptable. DA is run with inputs A, P, and q and student receive offers.

At Stage 2, students take achievement tests, with scores denoted Y_i.

2.3.1 Model Setup: stage 1

2.3.1.1 Priorities

One common objective of magnet re-assignment is to promote diversity in magnet schools, which are often located in areas with high concentration of poverty and low academic achievement. It is thus reasonable to assume that school districts design priority structures to give priorities to high-performing students from more affluent areas. Schools cannot perfectly measure student ability, so noisy priorities

will be granted. Student i's priority points at school s is defined by:

$$P_{is} = \theta a_i + (1-\theta)\xi_{is}, \theta \in [0,1] \tag{2.1}$$

where ξ_{is} is i.i.d. normally distributed random variables with mean zero and variance one for $i = 1,...,N$ and $s = 1,...,S$. θ is an indicator of the correlation between ability and priorities.

Schools may enforce strict priorities, meaning that students with higher priority points will always outrank those with lower points. Alternatively, in a more realistic setting, schools use priority categories by imposing the same priority for a group of students. For example, magnet schools in Wake County has priorities for students from high-performing areas in terms of academic achievement. As a result, multiple applicants may share the same priority at one school, which brings out random tie-breaking. This priority equation can be slightly manipulated to model the latter setting. We may grant the same priority to all students with priority points within a range, or specify a certain group of students to receive a priority, e.g. students in the top quartile of priority points.

Depending on the priority structure, student i will have priorities at schools where she has higher priority points, P_{is}. In this setting, school priorities is uncorrelated with student rank over schools. Although student ability, a_i, is a component in both U_{is} and P_{is}, it is unrelated to student preferences as it is the same across all schools for student i. However, a_i may be relevant in the comparison between U_{is} and U_{i0}, thus affecting the number of schools applied when $\omega \neq 1$.

2.3.1.2 Preferences

Students make application and enrollment decisions according the utility from each school. Preferences for schools may depend on expected academic achievement. Let Y_{i1} denote the potential test score for student i at any magnet school s. I model Y_{i1} following the school value-added model in Angrist et al. (2017) :

$$Y_{i1} = \mu_0 + \delta + a_i \tag{2.2}$$

where μ_0 is the mean potential outcome at the student's outside option school indexed 0. δ is the causal effect of magnet schools compared to the outside option, or traditional schools. a_i is student i's unobserved "ability", which is an i.i.d. normally distributed random variable with mean zero and variance σ_a^2 for $i = 1,...,N$. Alternatively, the potential test score at the outside option school, Y_{i0}, is modeled as:

$$Y_{i0} = \mu_0 + a_i \tag{2.3}$$

In addition to potential academic outcomes, students also consider school-specific characteristics, denoted by η_s, that are uncorrelated with Y_i in our model, such as diversity in this school and extra-curricular activities. Such characteristics contributes to the utility students generate from attending each magnet school, U_{is}. Another component of the utility model is an student- and school-specific

error term, v_{is}. v_{is} can be interpreted as noise when students measure η_s with imperfect information. An alternative explanation is that students value school characteristics differently. Student i's utility associated with attending school s is shown in the following equation:

$$\begin{aligned} U_{is} &= \alpha Y_{i1} + (1-\alpha)[\omega\eta_s + (1-\omega)v_{is}] \\ &= \alpha(\mu_0 + \delta + a_i) + (1-\alpha)[\omega\eta_s + (1-\omega)v_{is}] \end{aligned} \tag{2.4}$$

where η_s is an i.i.d. normally distributed random variable with mean zero and variance σ_η^2 for $s = 1, ..., S$. v_{is} is i.i.d. normally distributed random variables with mean zero and variance σ_v^2 for $i = 1, ..., N$ and $s = 1, ..., S$. Both α and ω are weights ranging between 0 and 1 that allows us to adjust the relative importance of each component of the utility model. Note that Y_{i1} does not affect student rank over schools as it is constant across all schools. Alternatively, v_{is} can be considered as individual's own preference over school characteristics.

For simplicity, we can assume that the utility from the outside option is measured without error:

$$\begin{aligned} U_{i0} &= \alpha Y_{i0} + (1-\alpha)\eta_{i0} \\ &= \alpha(\mu_0 + a_i) + (1-\alpha)\eta_{i0} \end{aligned} \tag{2.5}$$

Students compare U_{is} with U_{i0} all S schools and only apply to schools where $U_{is} > U_{i0}$, i.e. $\alpha\delta > (1-\alpha)\eta_{i0} - \omega\eta_s - (1-\omega)v_{is}$, with ranks following a descending order of U_{is}. Since Y_{i1} does not have any school-specific component, we can tone up or down the commonality of preference by changing ω. If $\omega = 1$, then every student will have the same submitted preference.

2.3.1.2.1 Correlated with priorities

An alternative way of defining student preference introduces correlation between preferences and priorities.

Students may believe they have a higher chance of receiving an offer at a magnet school where she has priorities. Thus, it may be reasonable to assume that she assigns higher utility points to this school. Student i's "adjusted utility points" at school s is defined by:

$$U'_{is} = \alpha Y_{i1} + (1-\alpha)[\omega\eta_s + (1-\omega)v_{is}] + \tau P_{is} \tag{2.6}$$

As with the basic preference setup, students will apply to schools with rankings based on utility points generated by each school. Students will not apply to schools with U'_{is} that is smaller than the utility from her outside option, U_{i0}.

At the end of stage 1, students apply to a list of magnet schools that gives higher utility points compared to their outside options. Schools also submit priorities over applicants. We conduct centralized assignment using the Deferred Acceptance (DA) mechanism and an assignment outcome for each student is generated. Some applicants receive a seat at one of the schools they applied to, while others do not have a seat at any magnet school.

2.3.2 Model setup: stage 2

In stage 2, students take tests and receive test scores, which are modeled as follows:

$$Y_i = \mu_0 + \delta D_i + a_i + X_i'\gamma + \epsilon_i \tag{2.7}$$

D_i is a binary variable that equals 1 if the student attends any magnet school and 0 otherwise. Recall that a_i is the unobserved student ability. X_i' is a vector of student characteristics that affects student outcome, such as gender and race. ϵ_i is idiosyncratic shock to outcome.

2.4 Propensity-score distribution metrics and assignment models

2.4.1 Metrics on ideal properties of distribution

The econometric method proposed by Abdulkadiroğlu et al. (2017a) is advantageous for two main reasons. First, it recovers the full range of assignment variation by including applicants subject to both oversubscribed and undersubscribed schools. The propensity score captures the variation created by applicants randomly assigned to undersubscribed schools when they are rejected by oversubscribed schools in lotteries. As a result, I consider propensity score distributions with higher density strictly between 0 and 1 as benefiting more from this novel identification method. Second, propensity score conditioning suffer much less from reduction in dimensionality compared to traditional methods that fully stratifies based on both preferences and priorities. For this reason, I am also interested in propensity score distributions with mass points that will provide more observations for each propensity score control.

Three measures are introduced to quantitatively evaluate each shape of propensity score distribution. A propensity score distribution with relatively high values in these measures will be less U-shaped, which is better for identification.

2.4.1.1 Weighted distance to mid-point

$WD(x) = \sqrt{2\sum_x (0.5 - |x - 0.5|)(\frac{freq(x)}{N})^2}$ This measure can take any value between 0 to 1. It increases in the density of propensity scores groups, which implies that more students are under the same propensity score. It also increases if the propensity score gets closer to 0.5, where lottery losers will have one-to-one matches with lottery winners. It is equal to the highest value, 1, when all propensity scores are equal to 0.5. This is the scenario with the highest number of propensity scores between 0 and 1. All students can be evaluated together because they have the same propensity score. The ratio of lottery losers to lottery winners will also be one-to-one, ensuring a maximum number of observations under each counterfactual.

2.4.1.2 Information Entropy

$H(x) = \sum_{i=1}^{N} [-PS_i \log_2 PS_i - (1-PS_i)\log_2(1-PS_i)]$

Information entropy is a measure of unpredictability of the average information content of an event. If an event is very probable, it is no surprise when it happens and thus provides little new information. Inversely, if the event was improbable, it is much more informative that the event happened. Propensity scores equal to zero have zero contribution to this measure. It increases in the uncertainty of assignment, which means higher number students have a propensity score strictly between 0 and 1.

2.4.2 Simulation results on propensity-score distribution metrics and assignment features

2.4.3 Assignment features

In Section 2, I have presented characteristics of assignments that in theory will affect propensity-score distributions. The model setup shown in Section 3 allows me to fine-tune features of assignments by varying one parameter at a time. Table 2.1 shows propensity-score distribution metrics and assignment features.

The first four rows explore the commonality of preferences by varying ω in Equation 2.4. To recall, ω is a parameter ranging between 0 and 1 that imposed completely common preference among all applicants when it equals to 1. with completely common preferences, students submit the same ranking over schools to the centralized mechanism. Intuitively, it will increase competition in the first few rounds at oversubscribed schools. As a result, students who lose the lottery at oversubscribed schools will be randomly enrolled at undersubscribed schools that they applied to. The econometric method by Abdulkadiroğlu et al. (2017a) is designed to capture this random variation. Results in Table 2.1 confirms my intuition that substantially more students will be subject to lottery as preference commonality increases. As expected, the two distribution metrics proposed in section 4.1 are positively correlated with the number of applicants subject to lottery.

Figure 2.1 plots the four propensity score distributions shown in row 1 to 4. The distributions are far from the bell shape. Possibly explanations include the multi-stage nature of propensity score distributions, which makes it unlikely that the proportion of students sharing the same propensity scores increases towards one propensity score. Another explanation is high variation in student types consisting of student preferences and priorities. Because students are so different from each other, sharing the exact same propensity is rare.This type of propensity score distribution has been found to be common in real-life assignments (Abdulkadiroğlu et al., 2017a; Dur et al., 2020).

The next four rows shows simulation result by varying the number of equal-sized priority classes. If this number is equal to one, it means every applicants shares the same priority class, making everyone applying to the same school subject to the same lottery. If this number is equal to three, it means one third of the applicants have the high priority while one third have the low priority at each school. The rest do not have priorities. Here, I set θ to be 0.5, which means some degree of commonality in school priorities. Students with the high priority at one school will not necessarily have the high priority at other schools. Student preferences are set to be partly common. Not surprisingly, the first row generates the

highest amount of applicants subject to lottery. The two propensity-score distribution metrics increases as the number of equal-sized priority classes decreases, suggesting potential benefits for identification.

Both the number of priority classes and the priority structure are important for the amount of variation created. A priority structure with a larger proportion of students having the top priority generates very different assignment results from a priority structure with very few students having high priorities. For the row with the "inverted pyramid" priority structure, I impose a priority structure where over half of total number of applicants have the highest priority while limiting the proportion of students with lower priority levels. For the row with the "pyramid" priority structure, the priority structure is more realistic with less than 10 percent of student having the top priority at each school. The majority of applicants do not have any priorities. Again, priority points among schools are partly correlated with each other. It is possible that students with low ability have high priorities at one or more schools. This result follows my intuition that the proportion of students at a conditionally seated priority class directly contributes to random variation captured by propensity scores. A pyramid priority structure naturally results in schools having conditionally seated priority classes with fewer students.

Lastly, I simulate assignments with various levels of correlation between priorities and preferences. τ is from Equation 2.6 that controls how priority points at a school affect utility points generated from that school. When τ is high, students are more likely to apply to schools where they enjoy higher priorities. As expected, as the level of correlation increases, the assignment created less variation. It should be noted that correlation will have limited impact on propensity scores when ω is small, i.e. preferences are already diverse.

2.4.4 Highly probable and improbable events

When simulating assignments and calculating propensity scores, researchers may encounter propensity scores that are very close to 0 or 1.[1] Table 2.2 presents two simple assignment setups to provide some intuition behind these "extreme" propensity scores. In both setups, priority points are random assigned to students. Equal priority classes are then imposed based on priority points. In the case of completely common preferences, there are no highly probable and improbable events that result in "extreme" propensity scores. The lowest and highest propensity scores subject to school-choice experiments are in ranges between 0.1 and 0.8. Common preferences suggest that, in DA, all students propose to the same school in each round if they are not waitlisted. The number of schools in the assignment is denoted by S. As a result, the assignment takes only S rounds and students will be subject to only S lotteries. Propensity scores will be determined solely by the number of seats at each school and the number of applicants subject to lottery. In this setup, "extreme" propensity scores will only occur when school capacities are too low or the number of priority classes is too high.

The occurrence of highly probable events increases as I introduce more dispersed preferences. The number of students with propensity scores higher than 0.975 is the highest when there are no priority classes. The rare events of not being assigned for those with high propensity scores occur likely due

[1] Abdulkadiroğlu et al. (2017a) excludes applicants with a rounded simulated propensity score larger than 0.975 or smaller than 0.025 in their semiparametric model.

to "bad" draws of tie-breakers. The highly improbable events are reflected by propensity scores lower than 0.025. The number of students with propensity scores lower than 0.025 increases in the number of priority classes. These rare events of assignment occur likely due to competition for the limited number of seats subject to lottery at popular schools, since high-priority students will always be assigned . It is possible that low-priority students can only receive assignments in exceptionally "good" draws of tie-breakers.

The propensity scores presented in Table 2.2 indicate that the "extreme" propensity scores are not uncommon. Even simple setups with 50% common preferences and random priorities can generate both highly probable and improbable assignments documented by propensity scores higher than 0.975 or lower than 0.025. Part A and Part B also demonstrate that the distribution of highly improbable and probable events are not necessarily symmetrical. It is unlikely the case that the highly probable and improbable events of assignment happen in the same simulations with "weird" draws of tie-breakers. Rather, it appears that these events are due to dispersed competition at a set of popular schools. Part C presents propensity scores in Wake County's magnet assignment for the 2015-2016 school year. In real-life assignments, preferences and priority structures may be more complex, introducing additional "disrupting" factors that may lead to more frequent occurrences of propensity scores that are close to 0 or 1.

2.5 Propensity-score distribution metrics and estimation results

Taking simulation data on centralized assignments based on the last block in Table 2.1, I follow Equation 2.7 and create an outcome variable for each student. Components of the outcome variable includes the mean potential outcome at the student's outside option, which can be a traditional public school, a school outside of the public school district. The equation contains a binary variable indicating whether a student has received a seat at any magnet school. The assignment is randomly selected from the assignment simulation results. I impose true causal effect of magnet school attendance. Students' unobserved ability, an observed characteristic are also in the equation, and an error term are also included in the outcome variable. [2] In this paper, I assume perfect enrollment to magnet schools and focus on the reduced-form estimation shown in Abdulkadiroğlu et al. (2017a).

I conduct Monte Carlo simulations to estimate the causal effect of magnet school, δ. In each round, all the parameters are set except for the error term, which is i.i.d. normally distributed. I run OLS regression and the propensity-score conditioning regression by Abdulkadiroğlu et al. (2017a) (AANP) of the outcome variable on the offer indicator and the single observed variable. I then record the estimation results by both models. The simulations are run 10,000 times for each assignment setup and each estimation model.

Table 2.3 shows the estimation results for each propensity score distribution when the true causal effect of magnet schools is 0.3. The third and fourth columns are calculated by averaging 10,000 OLS and

[2] For the observed characteristic, I generated a random binary variable that equals 1 for half of the applicants and 0 for the rest. This can be interpreted as the gender variable. I assign a positive parameter to the gender variable, since multiple studies have shown differential gain to charter schools and magnet schools by gender (Deming et al., 2014; **?**).

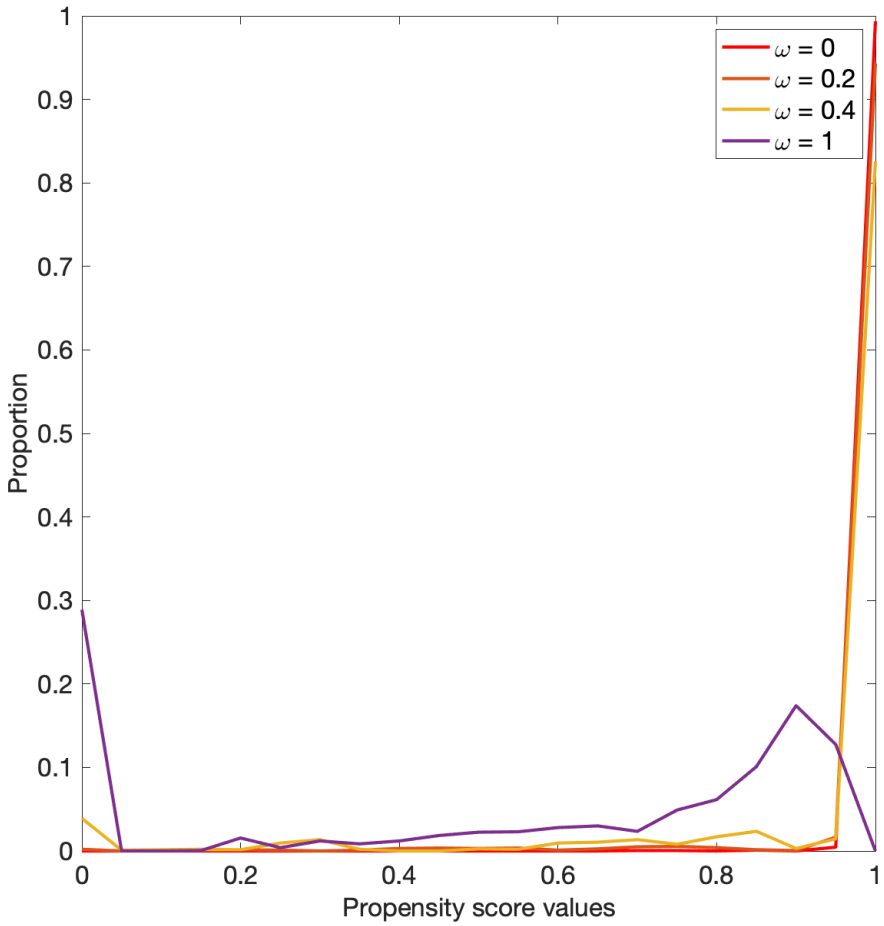
AANP estimators of δ. As can be seen from the table, the OLS substantially overestimates the magnet school effects. This is because ability is an unobserved component of the outcome that is positively correlated with priorities, and thus with the chance of receiving a magnet school offer. The AANP estimator is most precise when τ is between 0.4 and 0.6. The OLS results in substantial overestimation of the treatment effect, thus the risk of falsely not rejecting the null hypothesis of a null treatment effect is zero. The AANP method results in slightly higher Type II error rates.

Table 2.4 follows the same setup as in Table 2.3 with the true value switched to zero. With the null specification, I focus on the empirical Type I error rates generated by both methods. The OLS overestimates the magnet effect due to omitted variable bias. The AANP method mistakenly rejects the null hypothesis of zero effect at a rate around 4 percent given a 5% significance level. Similar to Table 2.3, the average estimator is more precise when there is some level of correlation between preferences and priorities.

2.6 Conclusion

This paper is one of the first to closely examine propensity scores as defined in Abdulkadiroğlu et al. (2017b), a method that improves on previous identification strategies by fully exploiting the random variation in centralized assignments. The propensity score is the probability of receiving an assigned seat. I generate propensity scores by manually constructing student preferences and school priorities and then conducting centralized assignments based on these data and randomly selected tie-breakers. In doing so, I examine the relationship between assignment features and propensity scores. I have also created a set of metrics of propensity scores that is linked to preferable characteristics in identification. By measuring the variation in estimation precision due to differences in propensity score distributions, this paper points out a unique way to valuate "data quality" through studying the preference/priority structures and assignment mechanisms.

Figure 2.1 Propensity score distribution with varying commonality of preference



Note: ω is a parameter ranging between 0 and 1 that imposed completely common preference among all applicants when it equals to 1. with completely common preferences, students submit the same ranking over schools to the centralized mechanism. Other model setups follows the first four rows in Table 2.1.

Table 2.1 Propensity-score distribution metrics by assignment features

Varying Parameters	Value	Model setup		Propensity score distribution		
		Preference	Priority	N of students	Information entropy	Weighted distance
ω (Random preference)	0	Part common utility points, $\omega = 0.8$	Partly common priority points, $\theta = 0.5$; 4 equal priority classes	35	6.170	0.001
	0.2			132	75.070	0.004
	0.4			305	199.466	0.006
(Common preference)	1			1422	902.856	0.020
Number of equal-sized priority classes	0			1982	1907.775	0.052
	1			1808	1507.129	0.036
	2			1367	1046.336	0.022
	5			619	491.700	0.012
Priority structures	Inverted pyramid	Part common utility points, $\omega = 0.8$	Partly common priority points, $\theta = 0.5$	1855	1612.804	0.038
	Pyramid			812	557.513	0.013
τ (Uncorrelated)	0	Part common utility points, $\omega = 0.8$, correlated with priority points	Partly common priority points, $\theta = 0.5$; 4 equal priority classes	1348	579.095	0.012
	0.2			1068	608.122	0.012
	0.4			1169	563.678	0.012
(Largely correlated)	1			599	326.083	0.008

Note: In this simulation, the following parameters are imposed: $\alpha = 1$, $\mu_0 = 0.1$, $\delta = 0.3$, $\theta = 0.5$, $\sigma_\eta = 0.5$, $\sigma_a = 0.5$, $\sigma_\nu = 0.5$, $\sigma_\xi = 0.5$.

Table 2.2 Two simple assignment setups and number of students with highly probable and highly improbable propensity scores

		Propensity score properties			
Number of priority classes	(0,1)	≤ 0.025	≥ 0.975	2nd smallest value	2nd largest value
Part A: 100% common preferences					
0	2,000	0	0	0.53	0.57
2	1,970	0	0	0.14	0.67
5	1,846	0	0	0.28	0.79
Part B: 50% common preferences					
0	1,504	0	267	0.44	0.995
2	980	1	226	0.02	0.9998
5	683	24	242	0.009	0.9998
Part C: Wake County 2015-2016 magnet assignment					
Unequal priority classes	468	129	16	0.0002	0.999999

Note: In this simulation, the following parameters are imposed: $\alpha = 1$, $\mu_0 = 0.1$, $\delta = 0.3$, $\theta = 0$, $\sigma_\eta = 0.5$, $\sigma_a = 0.5$, $\sigma_\nu = 0.5$, $\sigma_\xi = 0.5$. In Part A, 100% common preferences are assumed, $\omega = 1$. In Part B, 80% common preferences are assumed, $\omega = 0.5$.

Table 2.3 Propensity-score distribution metrics and estimation results (true value = 0.3)

τ	Information entropy	Weighted distance	Average $\hat{\delta}$ using OLS	Average $\hat{\delta}$ using AANP	Power (1-Type II error) using OLS	Power (1-Type II error) using AANP
0	514.974	0.012	0.905	0.277	1	0.999
0.2	494.259	0.012	0.949	0.322	1	1
0.4	468.951	0.011	0.962	0.295	1	1
0.6	435.626	0.01	0.969	0.307	1	1
0.8	398.561	0.009	0.972	0.279	1	0.998
1	354.265	0.009	0.979	0.264	1	0.993

Note: The assignment setup follows the last block in Table 2.1. Estimators and hypothesis testing statistics are recorded from 10,000 runs using the OLS and AANP models.

Table 2.4 Propensity-score distribution metrics and estimation results (true value = 0)

τ	Information entropy	Weighted distance	Average $\hat{\delta}$ using OLS	Average $\hat{\delta}$ using AANP	Empirical Type I error using OLS	Empirical Type I error using AANP
0	514.974	0.012	0.604	-0.025	1	0.041
0.2	494.259	0.012	0.648	0.02	1	0.04
0.4	468.951	0.011	0.662	-0.005	1	0.031
0.6	435.626	0.01	0.669	0.006	1	0.03
0.8	398.561	0.009	0.672	-0.021	1	0.043
1	354.265	0.009	0.679	-0.036	1	0.068

Note: The assignment setup follows the last block in Table 2.1. Estimators and hypothesis testing statistics are recorded from 10,000 runs using the OLS and AANP models.

CHAPTER

3

INCOME SECURITY AND FINANCIAL WELL-BEING OF OLDER WORKERS TRANSITIONING INTO RETIREMENT

3.1 Introduction

Income security is an important component of post-retirement well-being. Financial resources and security are found to be linked with increased mortality and poorer physical and mental health (Cutler et al., 2008; Diener and Chan, 2011; Kim et al., 2003). While the shift from defined-benefit pension plans to defined-contribution plans in the private sector places greater responsibilities on individuals to sufficiently plan and save for retirement, public workers are more likely to have defined-benefit pension plans that provide a steady stream of income in retirement (Munnell et al., 2007). In addition to pension benefits, the majority of North Carolina public retirees are also enrolled in the state health plan. However, a non-trivial proportion of public workers experience financial distress in retirement (Clark and Liu, 2020). This paper studies the income security of a sample of North Carolina public-sector employees as they enter retirement between 2014 and 2018, who receive pension benefits from a defined-benefit plan. I measure income security through examining how well pension benefits replace recent retirees' pre-retirement salaries and how their household income changes after retirement. I also explore the relationship between these income security measures and financial well-being indicators.

[0]I thank Robert Clark for his suggestions, guidance, and supervision on this project. I gratefully acknowledge funding by the Institute of Consumer Money Management. I also thank Melinda Morrill and Thayer Morrill for productive discussions.

Even though public retirees in this paper receive monthly pension benefits, wealth accumulated while working can still play an important role in retirement income security. Retirees may face difficult choices between wealth drawdown and reducing household expenditures if income from Social Security and pensions alone do not allow retirees to maintain their pre-retirement standard of living. Many without sufficient pre-retirement savings may suffer income shortfalls and resort to high-interest debt borrowing, which is linked with serious long-term financial consequences such as bankruptcy (Gross and Souleles, 2002).

Public workers typically have more financial security than their private sector counterparts, since they are more likely to have a defined-benefit pension plan that pays a steady stream of income throughout the retirement. Nonetheless, our previous work has documented that a nontrivial percentage of public workers have experienced financial distress in retirement in terms of holding credit card debt and not having sufficient emergency funds (Clark and Liu, 2020).

In this paper, I draw our analysis base on two waves of survey data merged with administrative records for North Carolina public employees. This paper focuses on individuals who transitioned into retirement between two of our survey waves. The other two samples include older workers who kept working and those who stayed retired during all of our surveys. I compare characteristics of these two samples with those who recently retired to better follow their income trajectory as they transition into retirement and understand changes in their financial well-being. I find that recent retirees have relatively high pension benefit replacement rates that are similar to Social Security benefit replacement rates. They are much more likely to experience large household income drops by the second survey, even after controlling for other demographic and financial characteristics.

My results are also benchmarked against to both public- and private-sector retirees in the Health and Retirement Study (HRS) dataset. The HRS is a rich and detailed nationally representative survey on older individuals over the age of 50. This study tracks employment, wealth, health, and well-being in older households both over time. While the North Carolina public employees have higher household income and are more likely to have a college degree, I find many similarities between the two samples in terms of income changes, financial literacy, and financial distress.

This paper begins with a discussion of the literature on income replacement and income security as older workers transition into retirement. In addition, literature on the financial distress and financial literacy of elderly households is reviewed. Section 3.3 describes our data sources and dataset construction. Section 3.4 presents descriptive statistics on income replacement of recent retirees while Section 3.5 examines the characteristics of subgroups experiencing large income changes. Finally, Section 3.6 shows regression analysis of the factors associated with income changes and studies the link between income changes and financial distress and Section 3.7 concludes. Appendix A and Appendix B offers description of our samples of North Carolina public workers and HRS respondents in greater detail. Appendix C includes tables on sample characteristics and regression analysis of additional outcomes. Appendix D records the survey questions I use to create key variables for our North Carolina dataset.

3.2 Literature review

3.2.1 Income replacement

There have been many studies on income security transitioning into retirement using both administrative data and household surveys over the past 30 years. Income replacement ratios, an approach to assess retirement income adequacy, evaluate the ratio of household income needed to sustain post-retirement consumption as a proportion of annual pre-retirement income.[1] While some have estimated total income replacement ratios, a number of studies have estimated the proportion of pre-retirement income replaced by Social Security benefits (Biggs and Springstead, 2008; Fox, 1979, 1982; Grad, 1990; Mitchell and Phillips, 2006). Replacement ratios are also calculated based on both Social Security benefits and pension income (Fox, 1982; Grad, 1990). Total income replacement ratios are computed based on the total income of a worker pre- and post-retirement, including any benefit payments from Social Security, employer pension system, as well as payout from individual retirement savings accounts. Researchers have found much variation of total income replacement ratios by age, cohort, and marital status. Butrica et al. (2012) estimated levels and sources of income at age 67 using the Social Security Agency's Modeling Income in the Near Term (MINT) model, which matches Social Security earnings records to results of the Census Bureau's Survey of Income and Program Participation. They showed that median price-indexed household earnings replacement ratios at age 67 is around 110 percent, which similar to Biggs and Springstead (2008)'s estimate of 106 percent using the same data. Smith (2003) used data from both the Current Population Survey and the Panel Study of Income Dynamics to estimate total income replacement ratios over the period 1977–1999. He estimated that the average pretax income replacement ratio at age 70 increased from 67 percent in 1977 to 74 percent in 1999. Munnell and Soto (2005) used data from the Health and Retirement Study (HRS) to estimate total replacement ratios and found that average replacement ratios of 79 percent for married couples and 89 percent for non-married individuals. Yuh (2011) analyzes data from the Federal Reserve's Survey of Consumer Finances and uses total retirement wealth of pre-retired households to project a median replacement ratio of about 67 percent of households' permanent income. Binswanger and Schunk (2012) use a specifically designed internet survey to elicit adequate levels of retirement spending. Their calculated minimum income replacement rates to finance spending is 95% for the lowest income quintile and 45% for the highest. Income replacement ratio calculated using benefit and earnings are closely related to policy considerations that design and update benefit systems. Biggs and Springstead (2008) used the MINT model and found the median Social Security replacement rate to be 47 percent. They showed the inverse correlation between lifetime earnings and replacement rate for both married and non-married individuals and households. Individuals with lower lifetime earnings receive significantly higher replacement rates under all measures than higher earning individuals. Similar to Butrica et al. (2012), they found that replacement rates are higher among older cohorts. Wu et al. (2013) used HRS data and show that median

[1] The replacement ratio approach largely ignores the utility theory, which results in fewer desirable theoretical properties compared to life cycle models. However, the replacement ratio is more robust against differences in modelling assumptions and is more widely used among retirement planning practitioners (Booth, 2004; Moore and Mitchell, 2000)

Social Security replacement rates vary by marital status, gender, and cohort, ranging between 40 and 53 percent. Poterba (2014) also found that Social Security as a source of retirement income is more important for those in the bottom half of the income distribution, which reflects both the low saving rate for many workers. They presented the Social Security benefit replacement rate to be 41 percent for someone with a national average wage. However, for an individual at the twenty-fifth percentile of the wage distribution, the replacement rate is more than 75 percent.

3.2.2 Financial distress

The majority of the financial distress literature primarily focuses on measuring the ability of households to remain current on incurred debts. Some studies examine financial distress using total household debt. Brown and Taylor (2008) examines "financial pressure" in Europe and the US through two measures, having negative net worth, and high debt-to-income ratios. Ampudia et al. (2016) proposed a comprehensive measure of financial distress that find households as being in distress if they are unable to pay their debts, i.e., do not have enough income to cover their spending and do not have sufficient liquid assets. They use information from the Household Finance and Consumption Survey (HFCS), a European household-level dataset. Athreya et al. (2019) exploits account-level panel data from the Federal Reserve Bank of New York Consumer Credit Panel/Equifax to study financial distress. Their definition of financial distress is having an account for which payment is at least 120 days past due in a given year. **?** examines self-evaluated over-indebtedness and found over a quarter of the sample overburdened with their debt loads.

Other research also studies non-housing short-term debts, as they potentially amplify the default risk for households. Christelis et al. (2009) uses data drawn from the first wave of the Survey of Health, Ageing and Retirement in Europe (SHARE) and examines financial fragility through both the objective measure of holding non-housing debt and the subjective measure of having difficulty making ends meet. Keys et al. (2020) measure financial distress using a nationally representative panel of TransUnion credit report data that tracks approximately 35 million individuals on a monthly basis over 2000-2016. They choose to focus on three metrics of financial distress – debt in collections, credit card non-payment, and personal bankruptcy.

There is also a strand of literature that focuses on subjective metrics that indicate financial distress not necessarily due to indebtedness.

Worthington (2006) takes an alternative approach that focuses on standards of living. Using data drawn from the Household Expenditure Survey in Australia, they define financial stress as being unable to have a holiday, to have meals with family and friends, and to engage in hobbies and other leisure activities, due to financial reasons. In order to study American workers' financial fragility in the wake of a financial crisis, Lusardi et al. (2011) examine respondents' perceived confidence to cope with an unexpected expense of $2,000 within the next month. Questions on "emergency funds" appears often in the financial planning literature, where planners often refer emergency funds as recommendation to households. They find that more than 35% of older individuals report not being able to cope with the unexpected expense. Brunetti et al. (2016) takes a novel approach and combine objective and subjective

measures. They define households as financially fragility if they earn sufficient income to at least cover all expected expenses, but that might be unable to cope with unexpected expenses. Their characterization focuses on non-optimal portfolio allocation with too small liquid assets using the Bank of Italy Survey on Household Income and Wealth (SHIW). They then use financial fragility to predict financial distress, which includes three indicators on having difficulty making ends meet, having informal debt, and having had arrears for 90 days or longer. In this paper, I focus on two type of financial distress, namely, having kept credit card balance and not being able to come up with $2,000 should need arises over the next month. I exploit changes in the latter variable between two surveys to show how income changes is associated with variation in financial distress.

3.2.3 Financial literacy

There has been much empirical research on measuring financial literacy of older workers using the "Big 3" questions designed by Lusardi and Mitchell (2011a) that were first introduced in a special module of the 2004 HRS survey. Since then, these three financial literacy questions have been widely used in several other U.S. surveys as well as related surveys around the world (Lusardi et al., 2010; Lusardi and Mitchell, 2011b, 2014).[2] The questions were created to measure basic knowledge of interest compounding, inflation, and risk diversification. The 2004 HRS survey results showed that only half could correctly answer a simple question on how two percent interest compounds over five years and the proportion of older individuals that gave correct answers to all three questions was only about one third (Lusardi and Mitchell, 2011a). Other studies have shown similarly widespread lack of these knowledge among older individuals in the U.S. Lusardi and Mitchell (2007b, 2014); **?**. Some sub-groups of the population, such as women and low-income individuals, have been found to be even less financially knowledgeable Lusardi and Mitchell (2007b, 2008); Seligman (2012). In addition, **?** showed that elderly individuals age 65 and above have the lowest debt literacy among all age groups despite having the highest self-assessed financial literacy.

Financial literacy has been found to be correlated with economic decisions with financial consequences of older workers. High levels of financial literacy are found to be positively associated with retirement planning, which is linked to more wealth accumulation (Lusardi and Mitchell, 2007b, 2011a,c). More financially literate individuals are also more likely to change saving goals for retirement, and to participate in the stock market (Clark et al., 2006; Stango and Zinman, 2009; van Rooij et al., 2011). Financial literacy also has consequences on income sources in retirement. Many older workers have limited knowledge of their pension and Social Security and these misconceptions can affect planned retirement timing and response to retirement incentives (Chan and Stevens, 2008; Clark et al., 2010, 2012).

As for household debt, several researchers have found links between low financial literacy and higher mortgage costs (Moore, 2003; Campbell, 2006). Stango and Zinman (2009) concluded that those unable to correctly calculate interest rates out of a stream of payments ended up borrowing more and

[2]For example, the 2007–2008 National Longitudinal Survey of Youth (NLSY) for young respondents (ages 23–28); the RAND American Life Panel (ALP) covering all ages; and the 2009 and 2012 National Financial Capability Study.

accumulating less wealth. Those with low financial literacy are also shown to be more likely to engage in costly borrowing behavior with higher transaction costs or higher fees for credit cards (Gathergood, 2012; **?**; Mottola, 2013). Allgood and Walstad (2013) found that both self-assessed and actual literacy are more likely to make in-time payments and less likely to carry credit card balance. In the long run, these high-cost borrowing accumulates and many financially illiterate households report being in excessive debts (**?**).

3.3 Data Description

The first part of the study examines the well-being of retired state and local government employees in North Carolina. I obtained administrative data files for all actively working employees covered by the North Carolina Teachers' and State Employees' Retirement System (TSERS) and the Local Governmental Employees' Retirement System (LGERS) as of March 2014. I extracted a stratified random sample who were sent two surveys developed by the authors and their colleagues in spring 2014, spring 2016, and spring 2018. I include public workers who were actively working as of our 2016 survey but have started claiming benefits by the 2018 survey.[3] The administrative records contain detailed information on each retiree including earnings, job information, years of service, creditable service, year of retirement, annuity option chosen, and benefit amount. The surveys obtained additional personal information not contained in the administrative data about race/ethnicity, education level, household income and wealth, work status after claiming retirement benefits and marital status, along with questions about their spouses of married retirees. The survey elicits self-reported household income in categories.[4] In addition, the survey included questions on retiree debt holding and financial literacy.

I divide the NC dataset into three subgroups. The first subgroup is the NC newly retired sample, which contains individuals who were still working as of our survey in 2014 or 2016 but have already started claiming benefits by our survey in 2016 or 2018. The NC newly retired sample represents workers who have transitioned into retirement between our surveys and is the focus of this study. Two other subgroups, serving as benchmarks for the NC newly retired sample, are the NC working sample and NC retiree sample, who were either both working or already retired in both our surveys in 2016 and 2018.[5]

In Clark and Liu (2020), I studied a sample of retirees who started claiming benefits between 2009 and 2014. Individuals in this paper are younger compared to those I studied in the past, with average age ranging between 58 and 62. It had only been over one year since initial benefit claiming for an average retiree. Other demographics are very similar to those shown in Clark and Liu (2020).[6]

For the NC dataset, the pension income replacement ratio is computed as dividing the annual

[3]The surveys were fielded as part of a grant from the Sloan Foundation examining the transition from career employment to complete retirement. For more information about the data and copies of the survey instrument, see: https://sites.google.com/site/publicsectorretirement/.

[4]Respondents were asked to report whether their annual household income fell into one of a series of brackets. Options were: less than $25,000, $25,000 to $49,999, $50,000 to $74,999, $75,000 to $99,999, $100,000 to $149,999, $150,000 to $249,999, $250,000 or more, and Don't Know. Individuals who responded 'Don't Know' when asked to report their annual income were excluded from the analysis.

[5]See Appendix B.1 and Appendix Table B.3 for more information on our surveys and sample construction.

[6]Appendix Table B.2 present mean demographics for all samples in this report.

pension benefit by the annual salary of 2013. The annual pension benefit of North Carolina retirees has a direct relationship to their employment history with state and local governments. The annual benefit formula is 1.82 or 1.85 percent of average final earnings times years of service, depending on whether the employee works for state or local governments. Thus, retirees with higher career earnings and more years of service will have higher annual pension benefits. Another component of the annual pension benefit is the choice to claim early with reduced benefit. Early retirement with reduced benefits is available to those who have reached age 50 and completed 20 years of creditable service and those who have reached age 60 and completed 5 years of service. The final determinant of pension benefit replacement ratio is the retiree's choice of annuity type. Upon termination and achieving the age and service requirements, retirees must request from the retirement system that their benefits begin and benefit amount varies based on their choice of annuity option.[7]

The second dataset draws on the Health and Retirement Study (HRS).[8] The Health and Retirement Study (HRS) is a longitudinal survey on older individuals that collects detailed individual-level information including demographics, employment, pension, and health of respondents and their spouses, as well as financial characteristics such as income and wealth at the household level. This study was originally designed to track age-eligible individuals and their spouses as they made the transition from active worker into retirement. Since the first wave in 1992, the HRS is an ongoing biennial survey. In addition to a repeated questionnaire, each wave contains a number of modules that provides even more detailed information on selected topics. I construct a sample for HRS data that is constrained to those who self-reported transitioning from to retirement at least once between 1994 and 2016, with valid pre- and post-retirement household income and responded to the 2010 wave. Since many individuals switch between work and retirement more than once, I focus on the income security of first-time retirees. Pre-retirement year is defined as the last wave when the individual reported employment status as working full-time or part-time before reported employment status as retired. Post-retirement year is recorded as year of the first wave after the post-retirement wave. For the HRS sample, the pension income replacement ratio is constructed through dividing the post-retirement year income from employer pension and annuity by the pre-retirement year household income.

3.4 Income changes of recent retirees

Table 3.1 lists the computed transition proportions for each baseline pre-retirement household income group for all four datasets. Both the NC newly retired sample and working sample are better off in household pre-retirement income compared to NC retirees, perhaps not surprisingly due to the fact that both samples were still working as of our first wave of survey in 2014 or 2016. Both samples have fewer workers in the lower two income groups than the NC retirees sample. 55-65% workers in these two samples were earning a household income of more than $75,000, compared to 43% in the NC retirees

[7] There are six annuity options, which include a single life annuity (called the maximum benefit), a 100 percent J&S, a 50 percent J&S, Social Security Leveling, and two additional J&S options with a pop-up provision if the retiree's beneficiary dies first.

[8] See Appendix B.2 for more information on the HRS.

sample and 44% in the HRS retirees sample. As the NC newly retired group progresses to the post-retirement survey, their income category composition becomes similar to that of the NC retiree group, with higher proportions of individuals in the lower three income categories and less than half remaining in the highest category. The HRS first-time retiree dataset contains the highest percentage of individuals earning low to moderate household income below $25,000 and between $25,000 and $50,000, which suggests our NC sample is relatively well-off when compared to a nationally representative dataset.

While the majority of all three NC samples maintain their household-income grouping two years in the future, there exists considerable variability. There are some similar trends shown in all panels. First, it is highly unlikely for the group with the lowest household income to transition to higher income categories, especially if the respondent has recently retired (Panel A, C, D, and E). Second, Panel A, D and E show that moderate-income households are struggling to maintain their income after the respondent transitions to retirement. Around 30% of households in the $25,000 to $50,000 group and 40% in the $50,000 to $75,000 group experience decreases in income categories, while individuals who continued to work or stay retired (Panel B and C) experience fewer changes in income. This indicates large income drops following retirement for both the NC newly retired and HRS samples. Finally, while the majority of the group with income higher than $75,000 stays in the same category, there is still 23-40% of retirees who slip two or three income categories post-retirement (Panel A, C, D, E). The high earners in the NC newly retired group has lower proportions of income decreases compared to the HRS sample. The HRS public-sector workers have higher income compared to private-sector workers, and are more likely to transition to a higher income category after retirement.

Table 3.2 shows the relationship between levels and changes of household income and pension income replacement rate for the retiree samples. In Table 3.2, I find that the median pension income replacement for both NC datasets is more than 40 percent.[9] Adding this to the Social Security benefit replacement rate of more than 40 percent shown in the literature, North Carolina retirees maintain a high total income replacement rate of more than 80 percent (Biggs and Springstead, 2008; Butrica et al., 2012; Wu et al., 2013). I find that pension income replacement rate is higher for households with higher pre-retirement income in both NC datasets, in contrary to the inverse relationship documented in the literature (Biggs and Springstead, 2008; Munnell and Soto, 2005; Smith, 2003). However, this is expected as the replacement rate largely represents total years of service and average salary, which is correlated with household income. For NC retirees, income category changes do not seem to be correlated with pension income replacement ratio. I observe income increases post-retirement positively correlated with pension replacement rates in the HRS sample. On the other hand, the median pension benefit replacement rate is highest among NC new retirees with no change in income categories, while the replacement rate among those with income increases and decreases are similar and lower at around 40%. This suggests that other sources of income, potentially income from work after retirement, also contribute to income category changes for NC new retirees. In all three panels, pension income replacement rate is higher among women, especially non-married women.

[9]The benefit formula for is 1.82 or 1.85 percent of average final earnings times years of service, depending on whether the employee works for state or local governments. However, the decision on early retirement with reduced benefits as well as choice over annuity types can also affect the pension income replacement rate.

Table 3.3 gives a closer look at the interaction of marital status and gender and corresponding income changes in each demographic subgroup. In all three panels for NC samples, non-married women are significantly more likely to be in the lower pre-retirement income categories than men and married women. Non-married women in the NC newly retired dataset (Panel A) are shown to be most well-off compared to other panels, yet less than a quarter earn household income higher than $75,000, compared to more than 70% among men and married women. Understandably, this is due to differences in the number of working household members. Since our surveys on North Carolina workers only elicits household income information in categories, I am unable to calculate household equivalence scale to directly compare non-married women with married households. Therefore, I need to conduct subgroup analyses to better understand the financial well-being of this subgroup. Women do not seem to experience more income changes than men post-retirement, nor are they more likely to have a pension replacement rate that is above median. For the NC working sample, it is worth noting that more married women have decreased income compared to men, while non-married women are more likely to have the same or higher income category in the second wave of survey two years later. Similar to the NC datasets, non-married women in the HRS sample have lower income categories and are more likely to experience income decreases compared to married women and men, despite the fact that they are slightly more likely to have positive income from employer pension benefits.

3.5 Income changes and wealth, financial literacy, and financial distress

Tables 3.4-3.7 present cross-tabulations for six subsamples - men, married women, and non-married women in NC newly retired dataset, NC working individuals and retirees samples, as well as the HRS sample. For each subsample, I highlight the relationship between income changes and components of individuals' finances, including wealth, financial distress, and financial literacy. Table 3.4 show the relationship between levels of household post-retirement Supplemental Retirement Plans (SRP) balance and income changes. The NC working sample and the retiree sample show similar distributions of SRP wealth, with the former group having slightly higher proportions of SRP coverage. There is notable difference in SRP wealth by marital status and gender for the newly retired group. Men are better off compared to women in terms of SRP balance and non-married women are far less likely to have SRP or have high SRP balance. Lower SRP coverage suggests that non-married women may lack wealth in retirement needed to buffer them against financial shocks.[10] Since I only have information on IRA and Keogh accounts for the HRS sample, more than half of the sample report having no such accounts. However, more than 20% of the sample still have a balance of over $100k, suggesting relatively high levels of retirement wealth.

While income changes do not seem to be correlated with the likelihood of having SRP or post-retirement SRP balance for the NC working sample and the retiree sample, a more complicated relation-

[10]Table B.4 shows that the newly retired sample are more likely to experience decreases in SRP wealth categories between two waves of survey compared to both the working sample and the retirees sample. Among the three subgroups of the newly retired sample, non-married women have the highest rate of SRP wealth decreases, suggesting both low levels and decreasing trend of wealth for this group of retirees.

ship is presented for the newly retired sample in Panel A-C, Table 3.4. For non-married women in the newly retired sample, those who experience decrease in income are worse off in terms of SRP wealth. Despite the small amount of observations, men and married women in the newly retired sample with income increases have lower levels of SRP wealth.[11] Having an above-median pension replacement rate is negatively correlated with SRP wealth for men and married women in the newly retired dataset, while a positive correlation is shown for non-married women in the newly retired dataset and retirees dataset. Non-married women in the newly retired sample with a below-median pension replacement rate are more likely to have decreases in SRP balance category, suggesting drawdown of retirement saving. For all other groups, having a below-pension replacement rate appears to be slightly negatively or not correlated with SRP balance decreases. The HRS sample draws stark contrast from the NC samples, with more than half of the sample with increases in IRA account balance, suggesting that these individuals are still accumulating wealth. Another explanation is that the changes in SRP balance in the HRS are not strong enough to trigger category changes as records in the NC surveys. The key takeaway from Table 3.4 and B.4 is that non-married women who are newly retired have the lowest levels of retirement savings while being most in need of supplemental saving from wealth drawdown. While non-married women having an above-median pension replacement rate is significantly less likely to experience SRP balance decreases, the relationship is reversed in other panels.

Table 3.5 presents additional sources of post-retirement income from work and Social Security, offering possible explanations to why some households have higher post-retirement income or SRP wealth. A large number of retiree households have supplemented their household income with post-retirement labor income, income from spouse, and Social Security benefit payments. Compared to the NC retiree dataset, the newly retired individuals are less likely to have worked for pay or have started claiming Social Security benefits. On the other hand, married new retirees are more likely to have a spouse that is still working, possibly due to this cohort's younger age on average. Perhaps unsurprisingly, I find increase in income categories positive correlated with both Social Security claiming and work after retirement. The only exception is non-married female new retirees, for whom I find higher proportions of Social Security claiming among those with decreases in income. However, non-married women are the subgroup with the lowest rate of work after retirement among the NC samples, which indicates that they are most in need of additional income in retirement. The HRS sample show similar proportions of individuals that have started claiming Social Security benefits.[12] However, they are less likely to have a spouse with employment.

Table B.5 and B.6 show the relationship between financial literacy and income changes.[13] Our surveys

[11] In all panels of Table B.4, those with income decreases are more likely to experience decrease in SRP wealth categories, possibly due to withdrawals from SRP accounts to supplement income. While the newly retired group may have accumulated wealth prior to retirement, it is less clear why 20% of the retirees group have increased SRP balance in two years. Explanations include contributions to SRP from work after retirement and respondents' recall errors on SRP balance.

[12] Since we define first-time retirees based on their self-reported employment status of "retired", no one in the HRS sample is considered having worked for pay after retirement.

[13] These three questions were designed by Lusardi and Mitchell and first implemented in the 2004 Health and Retirement Study (see Lusardi and Mitchell (2011a)). The same questions were subsequently added to several other surveys and have been used in numerous studies (Lusardi and Mitchell, 2014, 2011b, 2007a). While the questions seem rather simple, articles generally indicate a rather low level of correct answers across a wide range of sample populations.

asked three questions to measure public retirees' financial literacy. The objective of the questions was to determine if the retiree understood the power of compound interest, the impact of inflation of purchasing power, and the value of diversification in wealth management. These three simple questions have been used in numerous other studies to measure financial knowledge and to assess the importance of financial literacy in managing assets and debts. While financial literacy does not seem to be correlated with income changes or pension replacement rates, it is worth noting that the newly retired sample has relatively high levels of financial literacy compared to the working and retiree samples. Married women in the newly retired group, however, has the lowest average number of correct answers to financial literacy questions. Non-married women in the newly retired sample has the highest proportion experiencing decrease in financial literacy, suggesting many are unable to give a correct answer to questions they did well in two years earlier. In all samples, individuals with decreases in income are significantly more likely to have decreased financial literacy compared to those without changes in income categories. The HRS sample show similar rates of correct answers to financial literacy questions compared to the NC dataset. It is worth noting that the subgroup with below-median pension replacement rate are significantly more likely to give correct answers to all three questions.

Next, I turn to the consequences of income changes during the transition to retirement. I constructed two measures of financial distress. The first measure is defined as not being able to come up with $2,000 within the next month. This variable is available in S2016 and S2018, allowing us to investigate changes in financial distress over time. The second measure is defined as having kept a balance on credit cards as the household was unable to pay the full balance.

The second measure of having credit card debt is only available in S2018, making the financial distress index applicable to S2018 respondents. I choose the second measure of financial distress because it is comparable to the financial distress measure for the HRS sample, which is defined as having carried over unpaid balance on credit card last month. Table 3.6 shows the correlation between income category changes and financial distress. The newly retired sample is less likely to be financially distressed in either wave of survey compared to the other two samples. Yet, there is a downward trend in the second survey, with fewer individuals improving and more individuals becoming more distressed. Those with income category increases in the newly retired sample are also more likely to experience financial distress compared to active workers and retirees. Having an above-median pension replacement is positive correlated with financial distress with the exception of married women. Table 3.7 show financial distress for the HRS sample of public-sector new retirees. Public-sector non-married women is the subgroup with the highest proportion experiencing financial distress, while they are not as distressed as their counterparts in the NC newly retired sample. Public-sector men and married women are more financially distressed in the HRS sample compared to those in the NC newly retired sample, while non-married women have similar distress levels. As shown in Table 3.8, private-sector subgroups are in general less distressed compared to public-sector workers. In addition, there is a very little gap in financial distress levels by marital status and gender among private-sector workers in the HRS sample, contrasting that among NC and HRS public employees. Public-sector women with an post-retirement income increase are more likely to be distressed in both surveys and less likely to see their financial well-

being improve between the two waves of surveys. In the private sector, however, those with an income increase are not worse off than their peers. A potential explanation is that some public-sector workers receives higher income post-retirement due to seeking additional income sources such as working after claiming benefits when they are already financially distressed. On the other hand, private-sector workers' increase in post-retirement income is suggestive of their financial security.

3.6 Regression analysis of income changes and financial distress

Tables 3.9 and 3.10 are regression tables of two types of income changes – income (category) decrease.[14] For the NC sample, I find that compared to men, married women are more likely to have income category decreases and an above-median pension replacement rate. Non-married women in the HRS sample have higher rates of income decreases. Consistent with our expectations, those having recently transitioned to retirement are more likely to experience decrease in income categories, even after controlling for all covariates. Recent retirement does not seem to be linked with pension replacement rates. Having higher salary while working is also positive correlated with these two income change variables. Given the same retirement replacement rate, higher salary contributes to higher levels of change in income, which is more likely to be large enough to result in income category changes. The positive correlation between salary and pension replacement rate is possibly due to higher years of service contributing to both variables. Having a bachelor's degree is negatively associated with income decreases for both the NC and the HRS sample. There is little difference in the chance of experiencing income decreases by the sector of employment for the HRS respondents.

I examine the relationship between income changes and credit card debt in Tables 3.11 and 3.12. For the NC sample, increase in income category is positively correlated with financial distress, but the effect goes away after controlling for addition covariates including wealth and spousal characteristics.[15] This is likely due to its correlation with other demographic variables. For the HRS sample, having decreased income is statistically significantly associated with higher chance of having credit card debt. The result is robust even after controlling for IRA balance, financial literacy, and spousal characteristics. In all three regressions, non-married women are more financially distressed than men. Those with young age and low income are also more likely to experience financial distress in terms of having credit card debt. Having a high SRP balance is associated with lower chance of experiencing financial distress by almost 50% in all three regressions. Having a spouse with pension coverage or Social Security claiming brings an additional steady source of income, which negatively contributes to financial distress. Those with high self-rated financial knowledge has lower chance of carry credit card balance for the NC models. Objectively measured financial literacy, on the other hand, has a much smaller effect on distress in all three tables.

[14] Table B.7 present regression models of having an above-median pension replacement rate.

[15] Table B.8 shows that pension replacement rate does not seem to contribute to financial distress.

3.7 Conclusion

I use administrative data and survey responses on North Carolina public workers to examine changes in income near and after retirement. All workers in our dataset are enrolled in a defined-benefit pension plan and have started claiming benefits as of our second wave of survey. Compared to the nationally representative sample from the HRS, the NC datasets consist of workers and retirees with high pre-retirement household income as well as high pension income replacement ratios. Despite this, almost half of the newly retired sample has experienced decreases in post-retirement household income large enough to lead to income category decreases. I also find that many rely on additional sources of post-retirement income such as labor income directly or through a spouse. Moreover, the financial distress levels measured by having credit card debt is similar, if not higher, compared to the HRS sample. I find that public-sector workers in the HRS sample have similar characteristics to the NC samples in terms of demographics and financial literacy. Regression results show that for the NC sample and the HRS sample, high household income and high retirement savings account balance are strongly and robustly associated with less credit card debt. The HRS sample also show that those with income decreases transitioning into retirement are more likely to experience financial distress. The decision to retire, however, can be endogenous and correlated with events that lead to financial distress such as health shocks. In this case, negative life events could be driving retirement timing, income decreases, and increased levels of financial distress.

Public- and private-sector workers in the HRS sample have a few differences in the relationship between income changes and financial distress. Private-sector workers in the HRS sample have lower household income and similar levels of credit card holding compared to public-sector workers. One interesting observation is that the subgroup experiencing post-retirement income increases is the most distressed among public-sector workers while they are the least distressed among private-sector workers. Results suggest that the income increase is more likely to result from post-retirement work or Social Security claiming, rather than the drawdown of supplemental retirement savings. The relationship between income and financial well-being is possibly complicated by the prevalence of employer-provided define-benefit pension and work after benefit claiming behavior among public-sector workers. This result suggests that researchers and policy makers may look at additional measures on financial well-being measures and household consumption in addition to retirement income replacement for public workers.

Another distinct characteristic of the public-sector worker samples is the large gap by gender and marital status, consistent with the gender gap in retirement security documented by GAO (2020). Non-married women who recently retired fare the worst in almost all facets of their finance well-being, including household income, income changes, SRP wealth, early Social Security claiming, and access to emergency funds. The gap in financial distress between non-married women and other subgroups are even larger in the NC sample compared to the public-sector HRS sample. Despite having a median of 43% in pension income replacement rate, many non-married women rely on additional income sources by working after retirement and have already started claiming Social Security retirement benefits. Their

financial well-being may continue to deteriorate as they enter later phases of retirement and deplete their retirement savings. Policy makers may consider targeting non-married female retirees for financial education programs with the aim of promoting retirement planning and wealth accumulation during working years.

Table 3.1 Income category transition matrix

Pre-retirement household income	Proportion in sample	Post-retirement household income: Income ≤ $25K	$25K - $50K	$50K - $75K	> $75K
Panel A: NC newly retired					
Income < $25K	0.8%	50.0%	50.0%	0.0%	0.0%
$25K - $50K	12.5%	27.1%	57.6%	15.3%	0.0%
$50K - $75K	22.2%	3.8%	45.7%	41.9%	8.6%
> $75K	64.5%	0.7%	9.2%	20.0%	70.2%
N	473	24	112	114	223
Proportion		5.1%	23.7%	24.1%	47.1%
Panel B: NC working					
Income < $25K	1.3%	54.2%	25.0%	8.3%	12.5%
$25K - $50K	17.3%	1.5%	70.0%	23.2%	5.3%
$50K - $75K	25.8%	0.0%	5.6%	68.0%	26.4%
> $75K	55.6%	0.2%	1.1%	5.9%	92.9%
N	1,865	20	270	465	1,110
Proportion		1.1%	14.5%	24.9%	59.5%
Panel C: NC retiree					
Income < $25K	3.2%	66.7%	33.3%	0.0%	0.0%
$25K - $50K	25.9%	6.2%	61.5%	30.2%	2.1%
$50K - $75K	24.6%	0.0%	20.9%	51.6%	27.5%
> $75K	46.2%	0.0%	5.8%	17.0%	77.2%
N	370	14	92	105	159
Proportion		3.8%	24.9%	28.4%	43.0%
Panel D: HRS first-time retiree (public-sector)					
Income < $25K	15.0%	79.6%	14.6%	1.9%	3.9%
$25K - $50K	21.2%	37.1%	43.4%	10.5%	9.0%
$50K - $75K	19.5%	12.4%	23.9%	35.1%	28.6%
> $75K	44.3%	6.4%	9.5%	15.2%	68.9%
N	2,061	615	466	328	652
Proportion		29.8%	22.6%	15.9%	31.6%
Panel E: HRS first-time retiree (private sector)					
Income < $25K	30.3%	81.2%	14.3%	2.6%	1.9%
$25K - $50K	23.6%	37.3%	43.1%	11.8%	7.8%
$50K - $75K	15.8%	18.0%	29.0%	30.8%	22.3%
> $75K	30.3%	7.9%	12.9%	18.9%	60.3%
N	4,457	2,068	1,031	552	806
Proportion		46.4%	23.1%	12.4%	18.1%

Note: The NC samples only contain information on income categories. The HRS sample includes retirees who transitioned into retirement between two consecutive waves. HRS sample results are weighted.

Table 3.2 Income changes and pension income replacement rates

Respondent characteristics	Proportion in sample	Has non-zero pension benefit (HRS only)	Pension income replacement rate (percentile) 25th	50th	75th
Panel A: NC newly retired					
First-wave household income ≤ $50K	13.3%		0.22	0.34	0.50
$50K - $75K	22.2%		0.31	0.41	0.57
> $75K	64.5%		0.30	0.45	0.56
Income category increases	7.6%		0.30	0.41	0.61
Income category decreases	49.3%		0.28	0.40	0.54
Income category stays the same	43.1%		0.30	0.46	0.55
Men	34.7%		0.28	0.40	0.54
Married women	45.7%		0.31	0.44	0.55
Non-married women	19.7%		0.27	0.43	0.59
N	473		0.29	0.43	0.55
Panel B: NC retiree					
First-wave household income ≤ $50K	29.2%		0.21	0.36	0.47
$50K - $75K	24.6%		0.25	0.42	0.55
> $75K	46.2%		0.29	0.41	0.52
Income category increases	24.1%		0.25	0.41	0.49
Income category decreases	24.6%		0.27	0.40	0.52
Income category stays the same	51.4%		0.24	0.40	0.52
Men	35.7%		0.24	0.39	0.50
Married women	43.0%		0.27	0.41	0.52
Non-married women	21.4%		0.26	0.41	0.49
N	370		0.26	0.40	0.51
Panel C: HRS first-time retiree (public-sector)					
First-wave household income ≤ $50K	36.2%	32.6%	0.12	0.26	0.63
$50K - $75K	19.5%	44.0%	0.14	0.27	0.48
> $75K	44.3%	41.7%	0.11	0.26	0.48
Income increases	36.9%	40.9%	0.15	0.34	0.69
Income decreases	63.1%	37.6%	0.10	0.24	0.44
Men	48.0%	39.7%	0.13	0.26	0.54
Married women	21.8%	28.9%	0.10	0.24	0.45
Non-married women	30.3%	44.5%	0.12	0.28	0.51
N	2,061	38.8%	0.12	0.26	0.50

Note: For the NC samples, pension income is calculated based on monthly net benefit paid out retirees that are recorded in administrative records. For the HRS sample, pension income is self-reported income from employer pension or annuity. HRS sample results are weighted.

Table 3.3 Income changes and the interaction between marital status and gender

		Subsamples by marital status and gender		
Respondent characteristics	Proportion in sample	Men	Married women	Non-married women
Panel A: NC newly retired				
First-wave household income ≤ $25K	0.8%	0.6%	0.5%	2.2%
$25K - $50K	12.5%	7.3%	6.9%	34.4%***
$50K - $75K	22.2%	19.5%	17.1%	38.7%***
≥ $75K	64.5%	72.6%	75.5%	24.7%***
Income category increases	7.6%	5.5%	7.9%	10.8%
Income category decreases	49.3%	47.6%	51.4%	47.3%
Income category stays the same	43.1%	47.0%	40.7%	41.9%
Pension replacement below median	47.7%	52.1%	43.8%	48.8%
Pension replacement above median	52.3%	47.9%	56.2%	51.2%
Panel B: NC retiree				
First-wave household income ≤ $25K	1.3%	0.5%	0.7%	3.4%***
$25K - $50K	17.3%	9.4%	9.5%	42.7%***
$50K - $75K	25.8%	22.6%	22.0%	37.3%***
≥ $75K	55.6%	67.5%	67.8%	16.6%***
Income category increases	22.5%	23.1%	24.4%	18.0%**
Income category decreases	10.4%	8.0%	14.7%***	5.2%*
Income category stays the same	67.1%	68.8%	60.9%***	76.8%***
Panel C: NC Retirees				
First-wave household income ≤ $25K	3.2%	1.5%	1.3%	10.1%***
$25K - $50K	25.9%	18.9%	16.4%	57.0%***
$50K - $75K	24.6%	26.5%	22.6%	25.3%
≥ $75K	46.2%	53.0%	59.7%	7.6%***
Income category increases	24.1%	26.5%	23.9%	20.3%
Income category decreases	24.6%	23.5%	28.3%	19.0%
Income category stays the same	51.4%	50.0%	47.8%	60.8%
Pension replacement below median	53.0%	56.8%	50.3%	51.9%
Pension replacement above median	47.0%	43.2%	49.7%	48.1%
Panel D: HRS first-time retiree (public-sector)				
First-wave household income ≤ $25K	15.0%	9.5%	6.7%**	29.7%***
$25K - $50K	21.2%	15.0%	18.2%	33.4%***
$50K - $75K	19.5%	21.6%	17.7%*	17.5%
≥ $75K	44.3%	54.0%	57.3%	19.5%***
Income increases	36.9%	38.0%	37.5%	34.6%
Income decreases	63.1%	62.0%	62.5%	65.4%
Pension benefit is zero	61.2%	60.3%	71.1%***	55.5%
Pension replacement below median	13.4%	34.7%	34.2%	34.0%
Pension replacement above median	25.4%	65.3%	65.8%	66.0%

Note: Stars indicate proportions of individuals with income categories/changes in each group is significantly difference from proportions of those in the male workers group. HRS sample results are weighted. * p<.10, ** p<.05, *** p<.01.

Table 3.4 Income changes and household post-retirement SRP balance

Income changes	Proportion in sample	Post-retirement household SRP balance (IRA and Keogh only for HRS retirees)				
		No SRP	≤ $25K	$25K - $100K	≥ $100K	Don't know or Missing
Panel A: Men (NC newly retired)						
Income category increases	5.5%	0.0%	22.2%	22.2%	44.4%	11.1%
Income category decreases	47.6%	5.1%	16.7%	11.5%	65.4%	1.3%
Income category stays the same	47.0%	1.3%	14.3%	20.8%	61.0%	2.6%
Pension replacement ≤ median	52.1%	2.7%	17.6%	21.6%	56.8%	1.4%
Pension replacement > median	47.9%	2.9%	11.8%	13.2%	69.1%	2.9%
Panel B: Married women (NC newly retired)						
Income category increases	7.9%	23.5%	17.6%	5.9%	52.9%	0.0%
Income category decreases	51.4%	4.5%	13.5%	26.1%	51.4%	4.5%
Income category stays the same	40.7%	5.7%	17.2%	25.3%	46.0%	6.8%
Pension replacement ≤ median	43.8%	9.9%	13.6%	24.7%	44.4%	7.4%
Pension replacement > median	56.2%	5.8%	16.5%	26.2%	47.6%	4.8%
Panel C: Non-married women (NC newly retired)						
Income category increases	10.8%	10.0%	20.0%	30.0%	30.0%	10.0%
Income category decreases	47.3%	20.5%	27.3%	29.5%	22.7%	0.0%
Income category stays the same	41.9%	10.3%	20.5%	30.8%	33.3%	5.1%
Pension replacement ≤ median	48.8%	15.4%	12.8%	30.8%	33.3%	7.7%
Pension replacement > median	51.2%	19.5%	31.7%	26.8%	22.0%	0.0%
Panel D: NC working						
Income category increases	22.5%	5.5%	12.7%	24.2%	50.0%	7.9%
Income category decreases	10.4%	4.1%	15.5%	23.8%	52.3%	4.6%
Income category stays the same	67.1%	5.0%	13.2%	27.6%	46.6%	7.7%
Panel E: NC retirees						
Income category increases	24.1%	8.0%	12.5%	22.7%	51.1%	6.7%
Income category decreases	24.6%	8.8%	17.6%	19.8%	46.2%	7.7%
Income category stays the same	51.4%	10.6%	14.3%	19.6%	49.2%	6.8%
Pension replacement ≤ median	53.0%	8.8%	11.9%	15.5%	56.7%	8.2%
Pension replacement > median	47.0%	10.3%	17.8%	25.9%	40.2%	5.7%
Panel F: HRS first-time retiree (public-sector)						
Income increases	36.9%	48.4%	9.9%	15.9%	25.8%	-
Income decreases	63.1%	44.4%	13.4%	15.4%	26.8%	-
Pension benefit is zero	61.2%	46.6%	12.8%	14.7%	25.8%	-
Pension replacement ≤ median	13.4%	46.0%	11.0%	14.9%	28.2%	-
Pension replacement > median	25.4%	44.1%	10.9%	18.2%	26.8%	-

Note: The NC samples' SRP balance includes balance of any retirement savings plans (e.g., 401(k), 457, 403(b), or IRA). HRS sample only has information on IRA, Keogh accounts. SRP balance is reported in categories only for the NC samples. HRS sample results are weighted.

Table 3.5 Income changes and additional sources of income after retirement

		Return to work and Social Security (SS) claiming			
		Own		Spouse (married sample)	
Income changes	Proportion in sample	Has worked for pay after retirement	Has started claiming SS benefits	Has spouse that is currently employed	Spouse has started claiming SS benefits
Panel A: Men (NC newly retired)					
Income category increases	5.5%	55.6%	55.6%	57.1%	28.6%
Income category decreases	47.6%	38.5%	35.9%	39.1%	28.1%
Income category stays the same	47.0%	50.6%	44.2%	56.5%	26.1%
Pension replacement ≤ median	52.1%	43.2%	60.8%	52.5%	37.7%
Pension replacement > median	47.9%	52.9%	22.1%	50.9%	17.5%
N	164	45.1%	40.9%	48.6%	27.1%
Panel B: Married women (NC newly retired)					
Income category increases	7.9%	52.9%	52.9%	47.1%	76.5%
Income category decreases	51.4%	44.1%	42.3%	44.3%	44.3%
Income category stays the same	40.7%	40.9%	37.5%	61.4%	37.3%
Pension replacement ≤ median	43.8%	39.5%	60.5%	36.4%	48.1%
Pension replacement > median	56.2%	49.0%	27.9%	59.0%	45.0%
N	216	43.5%	41.2%	51.5%	44.2%
Panel C: Non-married women (NC newly retired)					
Income category increases	10.8%	70.0%	40.0%		
Income category decreases	47.3%	31.8%	59.1%		
Income category stays the same	41.9%	43.6%	41.0%		
Pension replacement ≤ median	48.8%	41.0%	74.4%		
Pension replacement > median	51.2%	41.5%	29.3%		
N	93	40.9%	49.5%		
Panel D: NC Retirees					
Income category increases	24.1%	60.7%	62.9%	47.7%	52.3%
Income category decreases	24.6%	57.1%	52.7%	40.8%	46.5%
Income category stays the same	51.4%	45.3%	61.1%	39.4%	53.5%
Pension replacement ≤ median	53.0%	46.4%	73.0%	33.1%	64.1%
Pension replacement > median	47.0%	58.0%	44.3%	52.1%	36.4%
N	370	51.9%	59.5%	41.8%	51.3%
Panel E: HRS first-time retiree (public-sector)					
Income increases	36.9%	-	45.1%	40.9%	33.7%
Income decreases	63.1%	-	51.3%	36.7%	36.0%
Pension benefit is zero	61.2%	-	47.5%	40.3%	36.8%
Pension replacement ≤ median	13.4%	-	58.0%	34.9%	37.9%
Pension replacement > median	25.4%	-	48.0%	34.6%	29.1%
N	2,061	-	49.0%	38.3%	35.2%

Note: Spouse's working status and social security claiming is reported by respondent in NC samples. Retirement status in the HRS sample is based on employment status changing to retired or partly retired, thus not recording work after benefit claiming behaviors. HRS sample results are weighted.

Table 3.6 Income changes and financial distress (NC Sample)

	Lacks access to emergency funds				Has kept credit card balance
Income changes	Financial distress improves	Financial distress worsens	Distressed in neither survey	Distressed in both surveys	Distressed in second wave
Panel A: Men (NC newly retired)					
Income category increases	0.0%	0.0%	100.0%	0.0%	0.0%
Income category decreases	0.0%	5.9%	90.2%	3.9%	12.0%
Income category stays the same	6.2%	10.4%	83.3%	0.0%	22.9%
Pension replacement ≤ median	2.5%	5.0%	92.5%	0.0%	15.0%
Pension replacement > median	2.3%	9.3%	86.0%	2.3%	19.0%
N	2.9%	7.8%	87.3%	2.0%	16.8%
Panel B: Married women (NC newly retired)					
Income category increases	11.1%	0.0%	77.8%	11.1%	33.3%
Income category decreases	2.6%	5.3%	90.8%	1.3%	24.7%
Income category stays the same	1.7%	8.6%	84.5%	5.2%	22.4%
Pension replacement ≤ median	0.0%	4.4%	91.1%	4.4%	28.9%
Pension replacement > median	4.3%	4.3%	87.1%	4.3%	19.4%
N	2.8%	6.3%	87.4%	3.5%	24.3%
Panel C: Non-married women (NC newly retired)					
Income category increases	0.0%	0.0%	33.3%	33.3%	50.0%
Income category decreases	3.0%	12.1%	78.8%	6.1%	30.3%
Income category stays the same	8.0%	0.0%	84.0%	8.0%	32.0%
Pension replacement ≤ median	4.8%	9.5%	81.0%	0.0%	20.0%
Pension replacement > median	3.6%	7.1%	75.0%	14.3%	42.9%
N	4.9%	6.6%	78.7%	8.2%	31.7%
Panel D: NC Working					
Income category increases	9.8%	2.4%	82.0%	5.6%	33.5%
Income category decreases	3.4%	2.7%	86.5%	7.4%	32.7%
Income category stays the same	6.8%	3.2%	82.7%	6.9%	30.2%
N	7.2%	3.0%	82.9%	6.7%	31.2%
Panel E: NC retirees					
Income category increases	4.5%	5.6%	84.3%	4.5%	29.9%
Income category decreases	5.5%	7.7%	81.3%	5.5%	27.0%
Income category stays the same	3.7%	6.3%	84.2%	5.3%	25.9%
Pension replacement ≤ median	2.6%	4.6%	87.2%	5.1%	26.4%
Pension replacement > median	6.3%	8.6%	79.3%	5.2%	27.9%
N	4.3%	6.5%	83.5%	5.1%	27.1%

Note: Financial distress refers to not being able to come up with $2,000 within the next month and having kept credit card balance when spending exceeded income in the last year. This table does not contain those who transitioned into retirement or were working between S2014 and S2016, as S2014 does not contain information on financial wellbeing. Financial distress measured by access to emergency funds is available in both S2016 and S2018. Financial distress measured by keeping credit card balance is only available in S2018.

Table 3.7 Income changes and financial distress (HRS sample of public-sector workers)

Income changes	Distressed in second wave	Changes in financial distress levels after retirement: Financial distress improves	Financial distress worsens	Distressed in neither survey	Distressed in both surveys
Panel A: Men					
Income increases	22.7%	4.6%	8.4%	72.7%	14.3%
Income decreases	23.3%	10.0%	5.4%	66.7%	17.9%
Pension benefit is zero	22.0%	8.8%	3.6%	69.1%	18.4%
Pension replacement ≤ median	27.6%	5.2%	12.5%	67.2%	15.1%
Pension replacement > median	22.8%	7.4%	9.6%	69.9%	13.1%
N	23.1%	7.9%	6.6%	69.0%	16.5%
Panel B: Married women					
Income increases	35.4%	15.3%	11.3%	49.3%	24.1%
Income decreases	26.1%	22.5%	7.4%	51.4%	18.7%
Pension benefit is zero	28.2%	18.9%	9.2%	52.9%	19.0%
Pension replacement ≤ median	48.5%	0.0%	9.5%	51.5%	39.0%
Pension replacement > median	28.9%	27.9%	9.1%	43.1%	19.8%
N	30.4%	19.2%	9.2%	50.4%	21.2%
Panel C: Non-married women					
Income increases	30.3%	5.4%	7.5%	64.3%	22.8%
Income decreases	28.3%	20.4%	6.6%	51.3%	21.6%
Pension benefit is zero	25.1%	17.7%	5.3%	57.2%	19.7%
Pension replacement ≤ median	14.8%	18.1%	4.3%	67.1%	10.5%
Pension replacement > median	42.3%	10.3%	10.9%	47.5%	31.4%
N	28.9%	15.4%	6.9%	55.6%	22.0%

Note: Financial distress refers to having carried over credit card balance from last month to this month. This table does not contain those who retired before the 2010 wave, as the credit card debt question is only included in waves in and after 2008. All sample results are weighted.

Table 3.8 Income changes and financial distress (HRS sample of private-sector workers)

Income changes	Distressed in second wave	Changes in financial distress levels after retirement: Financial distress improves	Financial distress worsens	Distressed in neither survey	Distressed in both surveys
Panel A: Men					
Income increases	20.9%	7.3%	6.9%	71.8%	14.1%
Income decreases	24.9%	7.4%	11.2%	67.7%	13.6%
Pension benefit is zero	22.5%	8.0%	10.1%	69.6%	12.3%
Pension replacement ≤ median	16.7%	8.5%	7.2%	74.8%	9.5%
Pension replacement > median	37.3%	2.0%	8.2%	60.8%	29.1%
N	23.3%	7.4%	9.5%	69.3%	13.8%
Panel B: Married women					
Income increases	17.2%	13.4%	4.7%	69.4%	12.5%
Income decreases	32.9%	13.1%	11.1%	54.0%	21.8%
Pension benefit is zero	23.1%	12.2%	6.5%	64.7%	16.6%
Pension replacement ≤ median	38.8%	21.6%	16.3%	39.6%	22.4%
Pension replacement > median	34.6%	13.4%	14.7%	52.0%	19.8%
N	25.6%	13.2%	8.2%	61.1%	17.5%
Panel C: Non-married women					
Income increases	19.4%	9.9%	3.4%	70.7%	16.0%
Income decreases	24.9%	13.0%	10.2%	62.1%	14.7%
Pension benefit is zero	20.7%	12.6%	6.6%	66.7%	14.1%
Pension replacement ≤ median	46.7%	18.2%	20.5%	35.1%	26.2%
Pension replacement > median	19.0%	2.3%	5.5%	78.7%	13.5%
N	22.9%	11.9%	7.7%	65.3%	15.2%

Note: Financial distress refers to having carried over credit card balance from last month to this month. This table does not contain those who retired before the 2010 wave, as the credit card debt question is only included in waves in and after 2008. All sample results are weighted.

Table 3.9 Regression of income changes – income category decreases (NC sample)

Respondent / household characteristics in S2014/S2016	Income category decreases (1)	(2)	(3)	(4)
Married female	0.072*	0.063	0.080*	0.073*
	(0.040)	(0.041)	(0.042)	(0.042)
Non-married female	-0.010	-0.008	0.012	0.044
	(0.048)	(0.049)	(0.050)	(0.057)
African American	0.013	0.028	0.031	0.029
	(0.057)	(0.058)	(0.060)	(0.060)
Hispanic	0.037	0.002	-0.015	-0.019
	(0.166)	(0.162)	(0.160)	(0.161)
Bachelor's degree or above	-0.077*	-0.115**	-0.133***	-0.136***
	(0.043)	(0.046)	(0.047)	(0.047)
Age at survey	0.005	0.006	0.005	0.006
	(0.004)	(0.004)	(0.004)	(0.004)
Has good health	-0.103**	-0.099**	-0.062	-0.058
	(0.047)	(0.050)	(0.051)	(0.051)
Transitioned to retirement	0.255***	0.210***	0.209***	0.208***
between S2016 and S2018	(0.037)	(0.078)	(0.078)	(0.079)
Salary in 2nd quartile	0.144**	0.172***	0.182***	0.177***
	(0.064)	(0.065)	(0.066)	(0.066)
Salary in 3rd quartile	0.232***	0.244***	0.241***	0.235***
	(0.062)	(0.064)	(0.064)	(0.064)
Salary in 4th quartile	0.233***	0.280***	0.275***	0.273***
	(0.061)	(0.062)	(0.063)	(0.063)
Financial literacy questions all correct			0.034	0.033
			(0.038)	(0.038)
Financial literacy answers missing			0.141	0.154
			(0.130)	(0.131)
Self-rated financial knowledge is high			0.107***	0.107***
			(0.038)	(0.038)
Has supplemental retirement saving accounts			0.038	0.032
			(0.053)	(0.053)
Spouse currently employed				0.051
				(0.041)
Spouse covered by pension plan				0.026
				(0.041)
Control for agency type and retirement year	No	Yes	Yes	Yes
Sample size	843	843	843	843
Mean dependent variable	0.384	0.384	0.384	0.384

Note: The sample contains retired workers with non-missing records for net benefit paid out per month. Average marginal effects from probit models are reported. Salary is the 2013 annual salary based on administrative records. Financial literacy information in S2016 is used in models. * p<.10, ** p<.05, *** p<.01.

Table 3.10 Regression of income changes – income decreases (HRS sample)

Respondent / household	Income decreases			
characteristics in in pre-retirement wave	(1)	(2)	(3)	(4)
Married female	0.012	0.013	0.015	0.010
	(0.023)	(0.023)	(0.023)	(0.024)
Non-married female	0.061***	0.061***	0.060***	0.067***
	(0.020)	(0.020)	(0.020)	(0.021)
African American	-0.025	-0.024	-0.029	-0.028
	(0.021)	(0.021)	(0.021)	(0.021)
Hispanic	-0.003	-0.001	-0.005	-0.004
	(0.024)	(0.024)	(0.025)	(0.025)
Bachelor's degree or above	-0.031*	-0.027	-0.024	-0.024
	(0.017)	(0.017)	(0.018)	(0.018)
Age at survey	0.004***	0.003	0.002	0.003
	(0.001)	(0.002)	(0.002)	(0.002)
Has good health	-0.017	-0.020	-0.018	-0.018
	(0.019)	(0.020)	(0.020)	(0.020)
Public-sector worker	0.004	0.004	0.004	0.003
	(0.016)	(0.016)	(0.016)	(0.016)
Has positive salary in 1st quartile	0.032	0.030	0.030	0.030
	(0.021)	(0.022)	(0.023)	(0.023)
Has positive salary in 2nd quartile	0.148***	0.146***	0.148***	0.148***
	(0.020)	(0.022)	(0.022)	(0.022)
Has positive salary in 3rd quartile	0.200***	0.197***	0.198***	0.198***
	(0.022)	(0.024)	(0.024)	(0.024)
Has positive salary in 4th quartile	0.215***	0.211***	0.214***	0.213***
	(0.026)	(0.028)	(0.029)	(0.028)
Has IRA, Keogh accounts			0.022	0.024
			(0.017)	(0.016)
Financial literacy questions all			0.027	0.027
correct			(0.036)	(0.036)
Financial literacy is missing			-0.016	-0.017
			(0.026)	(0.026)
Spouse currently employed				0.007
				(0.022)
Spouse covered by pension plan				0.019
				(0.019)
Control for survey cohorts	No	Yes	Yes	Yes
Sample size	6,410	6,410	6,410	6,410
Mean dependent variable	0.615	0.615	0.615	0.615

Note: Average marginal effects from probit models are reported. Results are weighted. * p<.10, ** p<.05, *** p<.01.

Table 3.11 Regression of financial distress (credit card debt) on income category changes (NC sample)

Respondent / household characteristics	Kept credit card balance when spending exceeded income			
	(1)	(2)	(3)	(4)
Income category increases	0.072**	0.057	0.054	0.055
	(0.037)	(0.037)	(0.037)	(0.038)
Income category decreases	0.010	0.004	-0.000	-0.003
	(0.029)	(0.029)	(0.029)	(0.029)
Married female	0.056	0.051	0.041	0.042
	(0.046)	(0.046)	(0.046)	(0.047)
Non-married female	0.048**	0.043*	0.036	0.038
	(0.023)	(0.023)	(0.024)	(0.024)
African American	0.062**	0.033	0.039	0.035
	(0.032)	(0.031)	(0.032)	(0.032)
Hispanic	0.022	0.018	0.021	0.018
	(0.076)	(0.076)	(0.076)	(0.075)
Age at survey	-0.007***	-0.006**	-0.006**	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
Transitioned to retirement	-0.080*	-0.084*	-0.078*	-0.105**
between S2016 and S2018	(0.042)	(0.042)	(0.042)	(0.045)
Retiree in S2016 and S2018	-0.090	-0.081	-0.080	-0.109
	(0.097)	(0.098)	(0.098)	(0.096)
Household income $50-75k	-0.028	-0.014	-0.014	-0.015
	(0.029)	(0.030)	(0.030)	(0.030)
Household income ≥ $75k	-0.114***	-0.059*	-0.054*	-0.063**
	(0.029)	(0.030)	(0.030)	(0.032)
Has supplemental retirement saving		0.005	0.013	0.009
accounts		(0.029)	(0.029)	(0.029)
SRP balance ≥ $100k		-0.156***	-0.150***	-0.149***
		(0.022)	(0.022)	(0.022)
Financial literacy questions all			-0.004	-0.004
correct			(0.022)	(0.022)
Self-rated financial knowledge is			-0.089***	-0.086***
high			(0.023)	(0.024)
Spouse currently employed				0.051
				(0.044)
Spouse covered by pension plan				0.021
				(0.025)
Spouse started claiming SS benefits				-0.039
				(0.028)
Working after retirement				-0.011
				(0.040)
Has started claiming SS benefits				0.059
				(0.039)

Note: The sample contains both actively working and retired workers with a valid response to S2018. Sample size is 2,147. Average marginal effects from probit models are reported. Mean dependent variable is 0.294. Specifications also include bachelor's degree, years since retirement, self-rated health, and agency controls. * p<.10, ** p<.05, *** p<.01.

Table 3.12 Regression of financial distress (credit card debt) on income changes (HRS sample)

Respondent / household characteristics	Carried over unpaid credit card debt from last month to this month			
	(1)	(2)	(3)	(4)
Income decreases	0.056***	0.055***	0.049**	0.049**
	(0.019)	(0.020)	(0.020)	(0.020)
Married female	0.012	0.011	-0.001	-0.000
	(0.027)	(0.027)	(0.028)	(0.028)
Non-married female	0.026	0.025	0.042*	0.041*
	(0.019)	(0.019)	(0.023)	(0.023)
African American	-0.030	-0.031	-0.032	-0.032
	(0.032)	(0.032)	(0.031)	(0.032)
Hispanic	-0.035	-0.035	-0.036	-0.037
	(0.030)	(0.030)	(0.030)	(0.030)
Age at survey	-0.007**	-0.007**	-0.006**	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
Public-sector worker	0.017	0.018	0.018	0.018
	(0.021)	(0.021)	(0.021)	(0.021)
Income ≤$25k	0.082***	0.081***	0.071**	0.072**
	(0.030)	(0.030)	(0.030)	(0.030)
Income $50-75k	0.135***	0.135***	0.114***	0.114***
	(0.035)	(0.035)	(0.035)	(0.036)
Income ≥ $75k	0.128***	0.127***	0.093**	0.094***
	(0.033)	(0.033)	(0.036)	(0.035)
Has IRA, Keogh accounts	0.072***	0.072***	0.074***	0.074***
	(0.021)	(0.021)	(0.022)	(0.022)
IRA,Keogh balance ≥ $100k	-0.116***	-0.115***	-0.118***	-0.117***
	(0.033)	(0.032)	(0.033)	(0.032)
Financial literacy questions all correct		-0.017	-0.016	-0.016
		(0.040)	(0.041)	(0.041)
Financial literacy is missing		-0.028	-0.028	-0.028
		(0.030)	(0.030)	(0.030)
Spouse currently employed			0.027	0.026
			(0.030)	(0.030)
Spouse covered by pension plan			0.079***	0.079***
			(0.025)	(0.025)
Spouse started claiming SS benefits			-0.003	-0.005
			(0.030)	(0.030)
Has started claiming SS benefits				0.015
				(0.027)

Note: The sample contains HRS respondents who retired after the 2010 wave, as the credit card debt question is only included in waves in and after 2008. Sample size is 3,007. Average marginal effects from probit models are reported. Mean dependent variable is 0.248. Model specification also includes bachelor's degree, self-rated health, and retirement year. * p<.10, ** p<.05, *** p<.01.

BIBLIOGRAPHY

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak (2011). Accountability and Flexibility in Public Schools: Evidence from Boston's Charters And Pilots. *The Quarterly Journal of Economics 126*(2), 699–748.

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2017a). Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation. *Econometrica 85*(5), 1373–1432.

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2017b). Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation. *Econometrica 85*(5), 1373–1432.

Abdulkadiroğlu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters (2020). Do parents value school effectiveness? *American Economic Review 110*(5), 1502–39.

Abebe, G., M. Fafchamps, M. Koelle, and S. Quinn (2019). Learning management through matching: A field experiment using mechanism design. Technical report, National Bureau of Economic Research.

Alexander, K. L., D. R. Entwisle, and C. S. Horsey (1997). From First Grade Forward: Early Foundations of High School Dropout. *Sociology of Education 70*(2), 87–107.

Allgood, S. and W. Walstad (2013). Financial literacy and credit card behaviors: A cross-sectional analysis by age. *Numeracy 6*(2), 1–26.

Ampudia, M., H. van Vlokhoven, and D. Żochowski (2016). Financial fragility of euro area households. *Journal of Financial Stability 27*, 250–262.

Angrist, J. D., S. R. Cohodes, S. M. Dynarski, P. A. Pathak, and C. R. Walters (2016). Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics 34*(2), 275–318.

Angrist, J. D., P. A. Pathak, and C. R. Walters (2013). Explaining Charter School Effectiveness. *American Economic Journal: Applied Economics 5*(4), 1–27.

Athreya, K., J. Mustre-del Río, and J. M. Sánchez (2019). The Persistence of Financial Distress. *The Review of Financial Studies 32*(10), 3851–3883.

Barger, J. (2003). Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina end of grade reading assessment. North Carolina Teacher Academy Technical Report.

Biggs, A. G. and G. R. Springstead (2008). Alternate Measures of Replacement Rates for Social Security Benefits and Retirement Income. *Social Security Bulletin 68*, 1.

Binswanger, J. and D. Schunk (2012). What is an adequate standard of living during retirement? *Journal of Pension Economics and Finance 11*(2), 203–222.

Bohrnstedt, G. W. and B. M. Stecher (1999). Class Size Reduction in California 1996-1998: Early Findings Signal Promise and Concerns. Report Summary. Technical report, CSR Research Consortium.

Booth, L. (2004). Formulating retirement targets and the impact of time horizon on asset allocation. *Financial Services Review 13*(1), 1–18.

Borghans, L., B. Ter Weel, and B. A. Weinberg (2008). Interpersonal styles and labor market outcomes. *Journal of Human Resources 43*(4), 815–858.

Brocato, J. (1989). How much does coming to class matter? Some evidence of class attendance and grade performance. *Educational Research Quarterly 13*(3), 2–6.

Brown, S. and K. Taylor (2008). Household debt and financial assets: evidence from Germany, Great Britain and the USA. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 171*(3), 615–643.

Brunetti, M., E. Giarda, and C. Torricelli (2016). Is Financial Fragility a Matter of Illiquidity? An Appraisal for Italian Households. *Review of Income and Wealth 62*(4), 628–649.

Butrica, B. A., K. E. Smith, and H. M. Iams (2012). This is Not Your Parents' Retirement: Comparing Retirement Income across Generations. *Social Security Bulletin 72*, 37.

Callahan, M. P., D. S. Parker, J. R. Sherwin, and M. T. Anello (1999). *Evaluation of Energy Efficiency Improvements to Portable Classrooms in Florida*. Florida Solar Energy Center, 1679 Clearlake Rd.

Campbell, J. Y. (2006). Household finance. *The Journal of Finance 61*(4), 1553–1604.

Case, R., Y. Okamoto, S. Griffin, A. McKeough, C. Bleiker, B. Henderson, K. M. Stephenson, R. S. Siegler, and D. P. Keating (1996). The Role of Central Conceptual Structures in the Development of Children's Thought. *Monographs of the Society for Research in Child Development 61*(1/2), i–295.

Chan, S. and A. H. Stevens (2008). What you don't know can't help you: Pension knowledge and retirement decision-making. *The Review of Economics and Statistics 90*(2), 253–266.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics 126*(4), 1593–1660.

Christelis, D., T. Jappelli, O. Paccagnella, and G. Weber (2009). Income, wealth and financial fragility in Europe. *Journal of European Social Policy 19*(4), 359–376.

Clark, R. L., M. B. d'Ambrosio, A. A. McDermed, and K. Sawant (2006). Retirement plans and saving decisions: the role of information and education. *Journal of Pension Economics and Finance 5*(1), 45–67.

Clark, R. L. and S. Liu (2020). Financial Well-being of State and Local Government Retirees in North Carolina. In A. Lusardi and O. S. Mitchell (Eds.), *Remaking Retirement: Debt in an Aging Economy*. Oxford, UK: Oxford University Press.

Clark, R. L., M. S. Morrill, and S. G. Allen (2010). Employer-provided retirement planning programs. *Reorienting Retirement Risk management*, 36–64.

Clark, R. L., M. S. Morrill, and S. G. Allen (2012). The role of financial literacy in determining retirement plans. *Economic inquiry 50*(4), 851–866.

Cowan, R., C. Donlan, D.-L. Shepherd, R. Cole-Fletcher, M. Saxton, and J. Hurry (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology 103*(4), 786–803.

Cullen, J. B., B. A. Jacob, and S. Levitt (2006). The Effect of School Choice on Participants: Evidence from Randomized Lotteries. *Econometrica 74*(5), 1191–1230.

Cutler, D. M., A. Lleras-Muney, and T. Vogl (2008). Socioeconomic status and health: dimensions and mechanisms.

Deming, D. J. (2011). Better Schools, Less Crime? *Quarterly Journal of Economics 126*(4), 2063–2115.

Deming, D. J. (2014). Using School Choice Lotteries to Test Measures of School Effectiveness. *The American Economic Review 104*(5), 406–411.

Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014). School Choice, School Quality, and Postsecondary Attainment. *The American Economic Review 104*(3), 991–1013.

Diener, E. and M. Y. Chan (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being 3*(1), 1–43.

Dur, U., R. G. Hammond, M. Lenard, M. Morrill, T. Morrill, and C. Paeplow (2020). The Attraction of Magnet Schools: Evidence from Embedded Lotteries in School Assignment. Working Paper.

Engberg, J., D. Epple, J. Imbrogno, H. Sieg, and R. Zimmer (2014). Evaluating Education Programs That Have Lotteried Admission and Selective Attrition. *Journal of Labor Economics 32*(1), 27–63.

Finn, J. D. and C. M. Achilles (1990). Answers and Questions about Class Size: A Statewide Experiment. *American Educational Research Journal 27*(3), 557–577.

Fox, A. (1979). Earnings Replacement Rates of Retired Findings from the Retirement History Study. *Social Security Bulletin 42*, 17.

Fox, A. (1982). Earnings Replacement Rates and Total Income: Findings from the Retirement History Study. *Social Security Bulletin 45*, 3.

GAO (2020). Retirement Security: Older Women Report Facing a Financially Uncertain Future. Report to the Special Committee on Aging, U.S. Senate, U.S. Government Accountability Office.

Gathergood, J. (2012). Self-control, financial literacy and consumer over-indebtedness. *Journal of economic psychology 33*(3), 590–602.

Gersten, R., N. C. Jordan, and J. R. Flojo (2005). Early Identification and Interventions for Students With Mathematics Difficulties. *Journal of Learning Disabilities 38*(4), 293–304.

Good, R. H. and R. A. Kaminski (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly 11*(4), 326–336.

Good, R. H., D. C. Simmons, and E. J. Kame'enui (2001). The Importance and Decision-Making Utility of a Continuum of Fluency-Based Indicators of Foundational Reading Skills for Third-Grade High-Stakes Outcomes. *Scientific Studies of Reading 5*(3), 257–288.

Grad, S. (1990). Earnings Replacement Rates of New Retired Workers. *Social Security Bulletin 53*, 2.

Graves, J. (2010). The academic impact of multi-track year-round school calendars: A response to school overcrowding. *Journal of Urban Economics 67*(3), 378–391.

Green, M. and H. Doran (2000). Still No Room To Learn: Crowded NYC Schools Continue To Jeopardize Smaller Class Size Plans. A Follow-Up Report to" No Room To Learn" and to the Class Size Summit Working Papers. Technical report, Public Advocate for the City of New York.

Gross, D. B. and N. S. Souleles (2002). An empirical analysis of personal bankruptcy and delinquency. *The Review of Financial Studies 15*(1), 319–347.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2012). The Effect of School Choice on Intrinsic Motivation and Academic Outcomes. Working Paper 18324, National Bureau of Economic Research.

Heckman, J. J. and Y. Rubinstein (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review 91*(2), 145–149.

Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics 115*(4), 1239–1285.

Hoxby, C. M. and G. Weingarth (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Working Paper.

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Jepsen, C. and S. Rivkin (2009). Class Size Reduction and Student Achievement The Potential Tradeoff between Teacher Quality and Class Size. *Journal of Human Resources 44*(1), 223–250.

Jerrim, J. and A. Vignoles (2016). The link between East Asian 'mastery' teaching methods and English children's mathematics skills. *Economics of Education Review 50*, 29–44.

Jordan, N. C., D. Kaplan, M. N. Locuniak, and C. Ramineni (2007). Predicting First-Grade Math Achievement from Developmental Number Sense Trajectories. *Learning Disabilities Research & Practice 22*(1), 36–46.

Keys, B. J., N. Mahoney, and H. Yang (2020). What Determines Consumer Financial Distress? Place-and Person-Based Factors. Technical report, National Bureau of Economic Research.

Kim, J., E. T. Garman, and B. Sorhaindo (2003). Relationships among credit counseling clients' financial wellbeing, financial behaviors, financial stressor events, and health. *Journal of Financial Counseling and planning 14*(2).

Krohn, G. A. and C. M. O'Connor (2005). Student Effort and Performance over the Semester. *The Journal of Economic Education 36*(1), 3–28.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics 114*(2), 497–532.

Krueger, A. B. and D. M. Whitmore (2001). The Effect of Attending a Small Class in the Early Grades on College-test Taking and Middle School Test Results: Evidence from Project Star. *The Economic Journal 111*(468), 1–28.

Lin, T.-F. and J. Chen (2006). Cumulative class attendance and exam performance. *Applied Economics Letters 13*(14), 937–942.

Lindqvist, E. and R. Vestman (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics 3*(1), 101–28.

Lusardi, A. and O. S. Mitchell (2007a). Baby Boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of Monetary Economics 54*(1), 205–224.

Lusardi, A. and O. S. Mitchell (2007b). Financial Literacy and Retirement Planning: New Evidence from the Rand American Life Panel. SSRN Scholarly Paper ID 1095869, Social Science Research Network, Rochester, NY.

Lusardi, A. and O. S. Mitchell (2008). Planning and financial literacy: How do women fare? *American Economic Review 98*(2), 413–17.

Lusardi, A. and O. S. Mitchell (2011a). Financial Literacy and Planning: Implications for Retirement Wellbeing. In A. Lusardi and O. S. Mitchell (Eds.), *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*, pp. 17–49. Oxford, UK: Oxford University Press.

Lusardi, A. and O. S. Mitchell (2011b). Financial literacy around the world: an overview. *Journal of Pension Economics and Finance 10*(4), 497–508.

Lusardi, A. and O. S. Mitchell (2011c). The outlook for financial literacy. In A. Lusardi and O. S. Mitchell (Eds.), *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*, pp. 1–15. Oxford, UK: Oxford University Press.

Lusardi, A. and O. S. Mitchell (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature 52*(1), 5–44.

Lusardi, A., O. S. Mitchell, and V. Curto (2010). Financial literacy among the young. *Journal of consumer affairs 44*(2), 358–380.

Lusardi, A., D. J. Schneider, and P. Tufano (2011). Financially Fragile Households: Evidence and Implications. Working Paper 17072, National Bureau of Economic Research.

McMullen, S. C. and K. E. Rouse (2012). The Impact of Year-Round Schooling on Academic Achievement: Evidence from Mandatory School Calendar Conversions. *American Economic Journal: Economic Policy 4*(4), 230–252.

McMullen, S. C., K. E. Rouse, and J. Haan (2015). The distributional effects of the multi-track year-round calendar: a quantile regression approach. *Applied Economics Letters 22*(15), 1188–1192.

Mitchell, O. S. and J. W. R. Phillips (2006). Social Security Replacement Rates for Alternative Earnings Benchmarks. SSRN Scholarly Paper ID 1094839, Social Science Research Network, Rochester, NY.

Moore, D. L. (2003). Survey of financial literacy in Washington State: Knowledge, behavior, attitudes, and experiences. Technical report, Washington State Department of Financial Institutions.

Moore, J. F. and O. S. Mitchell (2000). Projected retirement wealth and savings adequacy in the health and retirement study. In *Forecasting Retirement Needs and Retirement Wealth*. Philadelphia: University of Pennsylvania Press.

Mottola, G. R. (2013). In our best interest: Women, financial literacy, and credit card behavior. *Numeracy 6*(2), 4.

Munnell, A. H., K. Haverstick, and M. Soto (2007). *Why have defined benefit plans survived in the public sector*. State and Local Pension Plans.

Munnell, A. H. and M. Soto (2005). What replacement rates do households actually experience in retirement? Technical report.

NCES (2000). Condition of America's Public School Facilities: 1999. Technical report, National Center for Education Statistics.

NCES (2007). Public School Principals Report on Their School Facilities: Fall 2005. Technical report, National Center for Education Statistics.

NCPS (2014). Accountability Brief - School Performance Grades. Technical report, North Carolina Public Schools.

NCPS (2015a). 4-Year Cohort Graduation Rate Report: 2012-13 Entering 9th Graders Graduating in 2015-16 or Earlier. Technical report, North Carolina Public Schools.

NCPS (2015b). North Carolina School Performance Grades. Technical report, North Carolina Public Schools.

NCPS (2016). Statewide Facility Needs Survey 2015-2016. Technical report, North Carolina Public Schools.

Parcel, T. L. and A. J. Taylor (2015). *The end of consensus: Diversity, neighborhoods, and the politics of public school assignments*. University of North Carolina Press.

Pears, K. C., H. K. Kim, P. A. Fisher, and K. Yoerger (2016). Increasing pre-kindergarten early literacy skills in children with developmental disabilities and delays. *Journal of School Psychology 57*, 15–27.

Poterba, J. M. (2014). Retirement Security in an Aging Population. *American Economic Review 104*(5), 1–30.

Ribner, A., E. Harvey, R. Gervais, and C. Fitzpatrick (2019). Explaining school entry math and reading achievement in Canadian children using the Opportunity-Propensity framework. *Learning and Instruction 59*, 65–75.

Rivera-Batiz, F. L. and L. Marti (1995). A School System at Risk: A Study of the Consequences of Overcrowding in New York City Public Schools. IUME Research Report 95-1.

Roehrig, A. D., Y. Petscher, S. M. Nettles, R. F. Hudson, and J. K. Torgesen (2008). Accuracy of the DIBELS Oral Reading Fluency Measure for Predicting Third Grade Reading Comprehension Outcomes. *Journal of School Psychology 46*(3), 343–366.

Romano, E., L. Babchishin, L. S. Pagani, and D. Kohen (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology 46*(5), 995–1007.

Rosenbaum, P. R. and D. B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika 70*(1), 41–55.

Seligman, J. (2012). Evidence on the financial capability of older workers facing lump-sum retirement plan distribution. *Accounting and Finance Research 1*, 177–195.

Shirley, M. (2017). The Effects of Overcrowding on Student Academic Performance in Kentucky High Schools. Working Paper.

Smith, J. (2003). Trends and Projections in Income Replacement during Retirement. *Journal of Labor Economics 21*(4), 755–781.

Smolkowski, K. and K. D. Cummings (2016). Evaluation of the DIBELS (Sixth Edition) Diagnostic System for the Selection of Native and Proficient English Speakers at Risk of Reading Difficulties. *Journal of Psychoeducational Assessment 34*(2), 103–118.

Stango, V. and J. Zinman (2009). Exponential growth bias and household finance. *The Journal of Finance 64*(6), 2807–2849.

van Rooij, M. C. J., A. Lusardi, and R. J. M. Alessie (2011). Financial literacy and retirement planning in the Netherlands. *Journal of Economic Psychology 32*(4), 593–608.

WCPSS (2015a). 2014-2015 Facilities Utilization. Technical report.

WCPSS (2015b). 2014-2015 Facilities Utilization. Technical report, Wake County Public School System.

WCPSS (2019a). 2013 Building Program: Our Budget. Technical report, Wake County Public School System.

WCPSS (2019b). Magnet Schools / Overview. https://www.wcpss.net/magnet.

Winters, M. A. (2020). Charter Schools in Newark: The Effect on Student Test Scores. Technical report, Manhattan Institute for Policy Research.

Worthington, A. C. (2006). Debt as a source of financial stress in Australian households. *International Journal of Consumer Studies 30*(1), 2–15.

WRAL (2018a). Cary parents speak out against school reassignment. *WRAL.com*. https://www.wral.com/cary-parents-speak-out-against-school-reassignment/17890831/.

WRAL (2018b). Cary parents unhappy about WCPSS reassignment plan that could move students across town. *WRAL.com*. https://www.wral.com/cary-parents-unhappy-about-wcpss-reassignment-plan-that-could-move-students-across-town/17805548/.

WRAL (2018c). Reassignment: Morrisville parents frustrated at switch to year-round school. *WRAL.com*. https://www.wral.com/reassignment-morrisville-parents-frustrated-at-switch-to-year-round-school/17956104/.

Wu, A. Y., N. S. Karamcheva, A. H. Munnell, and P. J. Purcell (2013). How Do Trends in Women's Labor Force Activity and Marriage Patterns Affect Social Security Replacement Rates? *Social Security Bulletin 73*(4), 1–24.

Yuh, Y. (2011). Assessing Adequacy of Retirement Income for U.S. Households: A Replacement Ratio Approach. *Geneva Papers on Risk and Insurance: Issues and Practice 36*(2), 304–323.

Zhang, H., C.-Y. Wee, J. S. Poh, Q. Wang, L. P. Shek, Y.-S. Chong, M. V. Fortier, M. J. Meaney, B. F. P. Broekman, and A. Qiu (2019). Fronto-parietal numerical networks in relation with early numeracy in young children. *Brain Structure and Function 224*(1), 263–275.

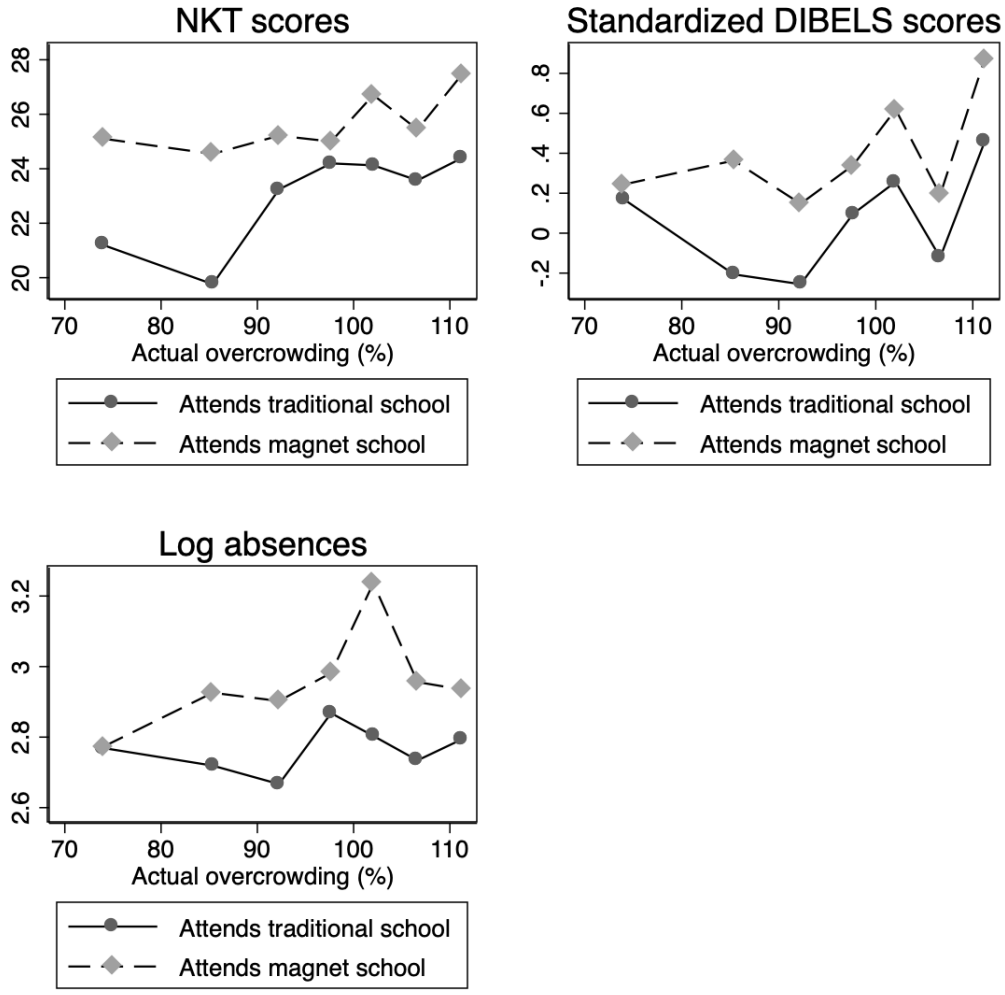
APPENDICES

APPENDIX

A

APPENDIX FOR CHAPTER 1

Figure A.1 Student outcomes by magnet attendance and base school overcrowding percentages



Note: Sample includes conditionally seated applicants who are recorded in the Wake County public school data continuously between the 2015-2016 school year and the 2018-2019 school year. Actual school overcrowding is calculated the actual enrollment over actual capacity in the 2014-2015 school year (WCPSS, 2015b).

Table A.1 Statistical tests for balance

Student characteristics	Non-offered, Non-overcrowded mean	Magnet offer * Overcrowded base	Magnet offer * Non-overcrowded base	Overcrowded base
	(1)	(2)	(3)	(4)
Base school is magnet	0.182	-0.004	-0.063	-0.130**
		(0.056)	(0.083)	(0.063)
Male	0.491	0.009	0.082	-0.020
		(0.075)	(0.110)	(0.083)
African American	0.270	0.128**	0.097	-0.078
		(0.059)	(0.087)	(0.066)
Hispanic	0.114	-0.000	-0.053	0.046
		(0.045)	(0.066)	(0.050)
Other race	0.114	-0.085	-0.108	0.254***
		(0.052)	(0.077)	(0.058)
Has sibling(s) enrolled in the district	0.211	0.035	-0.211**	0.107
		(0.067)	(0.099)	(0.074)
Has older sibling(s) enrolled in the district	0.128	0.058	-0.272***	0.188***
		(0.062)	(0.092)	(0.069)
Base school has below-median one-year teacher turnover rate	0.372	0.059	-0.049	0.451***
		(0.071)	(0.104)	(0.078)
Base school has below-median acts of violence	0.436	-0.038	-0.042	0.123
		(0.073)	(0.107)	(0.081)
Base school has below-median short-term suspensions	0.408	-0.030	0.073	0.242***
		(0.071)	(0.105)	(0.079)

Note: This table reports coefficients from regressions of student characteristics in each row on three variables shown in columns (2)-(4), as well as 9 propensity score controls indicating propensity score rounded to 0.1. The sample includes magnet school applicants who are recorded in the Wake County public school data continuously between the 2016-2017 school year and the 2018-2019 school year. * p<.10, ** p<.05, *** p<.01.

Table A.2 Applicant enrollments by base overcrowding status

Student enrollment type	All applicants subject to lottery	Lottery winner, over-crowded base school	Lottery winner, non-overcrowded base school	Lottery loser, over-crowded base school	Lottery loser, non-overcrowded base school
	(1)	(2)	(3)	(4)	(5)
Offered magnet school	0.129	0.776	0.688	0.000	0.000
Other magnet school	0.225	0.020	0.000	0.154	0.351
Non-magnet base school	0.420	0.102	0.042	0.687	0.365
Non-magnet non-base school	0.009	0.020	0.000	0.000	0.015
Non-magnet non-base school that is overcrowded	0.004	0.020	0.000	0.000	0.004
Non-magnet non-base chool that is non-overcrowded	0.005	0.000	0.000	0.000	0.011
Left WCPSS	0.215	0.082	0.250	0.159	0.269
Observations	550	49	48	182	271

Note: Proportions of applicants that enroll in each type of school in the 2015-2016 school year are shown in this table. Leaving WCPSS is defined as no school enrollment information in the Wake County Public School System administrative records.

Table A.3 First-stage results

	Magnet attendance (1)	Magnet attendance * Overcrowded base (2)
Received magnet offer * Overcrowded base	0.430***	-0.190***
	(0.061)	(0.045)
Received magnet offer * Non-overcrowded base	0.007	0.437***
	(0.020)	(0.048)
Overcrowded base	0.501***	-0.340***
	(0.054)	(0.028)
Base school is magnet	0.044***	0.054*
	(0.016)	(0.028)
Male	-0.010	-0.009
	(0.011)	(0.020)
African American	0.019	0.003
	(0.016)	(0.027)
Hispanic	0.015	0.069**
	(0.018)	(0.035)
Other race	-0.016	0.012
	(0.018)	(0.029)
Has sibling(s)	0.058**	-0.110***
	(0.023)	(0.039)
Not oldest	-0.072***	0.108**
	(0.025)	(0.043)
Below median acts of violence	0.029**	-0.005
	(0.014)	(0.022)
Below median short-term suspensions	-0.017	-0.074***
	(0.014)	(0.022)
Observations	973	973
First-stage F-statistics	27.53	216.14

Note: First-stage results for Table 1.3 is presented. Sample includes magnet school applicants who are recorded in the Wake County public school data continuously between the 2015-2016 school year and the 2018-2019 school year. Specifications include gender, race, siblings enrolled in the district, four base school characteristics, and nine propensity score controls indicating propensity score rounded to 0.1. Robust standard errors are reported. * p<.10, ** p<.05, *** p<.01.

Table A.4 Specifications checks for overcrowding effects - alternative definition for overcrowding

Student Outcomes	Actual Overcrowding ≥ 105%	Projected Overcrowding ≥ 105%	Mobile classroom percentage above median	Class size above district average
	(1)	(2)	(3)	(4)
Part A: Math				
NKT Score	-0.130	0.405	-0.595	-0.251
	(1.673)	(1.189)	(1.238)	(1.161)
NKT Not At-risk	-0.026	-0.025	-0.122*	0.001
	(0.094)	(0.069)	(0.070)	(0.066)
Part B: Literacy				
DIBELS Score	-0.118	0.324	0.444*	-0.540**
	(0.318)	(0.212)	(0.245)	(0.223)
DIBELS Not At-risk	-0.065	0.118**	0.046	-0.117*
	(0.092)	(0.059)	(0.069)	(0.067)
Part C: Non-cognitive				
Log Absences	0.544*	-0.176	0.043	0.050
	(0.289)	(0.169)	(0.178)	(0.177)

Note: This table presents four specification tests for Table 1.3. Specifications include gender, race, siblings enrolled in the district, four base school characteristics, and nine propensity score controls indicating propensity score rounded to 0.1. Robust standard errors are reported. * p<.10, ** p<.05, *** p<.01.

Table A.5 Specifications checks for overcrowding effects - alternative propensity score specification

Student Outcomes	PS is linear	PS rounded to 100th
	(1)	(2)
Part A: Math		
NKT Score	-0.093	0.029
	(1.590)	(1.570)
NKT Not At-risk	-0.018	0.015
	(0.092)	(0.092)
Part B: Literacy		
DIBELS Score	-0.108	-0.203
	(0.300)	(0.314)
DIBELS Not At-risk	-0.076	-0.085
	(0.087)	(0.088)
Part C: Non-cognitive		
Log Absences	0.367	0.498*
	(0.258)	(0.287)

Note: This table presents two specification tests for Table 1.3. Specifications include gender, race, siblings enrolled in the district, four base school characteristics, and propensity score controls. Robust standard errors are reported. * p<.10, ** p<.05, *** p<.01.

APPENDIX

B

APPENDIX FOR CHAPTER 3

B.1 NC sample construction and restrictions

This analysis of worklife transitions of public employees in North Carolina uses administrative records from the North Carolina Retirement Systems Division (RSD) and information from several surveys with responses merged with administrative records. The first and second survey are conducted among two cohorts of older workers who were employed full-time in 2014 and 2016, respectively. They are used to examine the retirement plans of older workers. The third and fourth survey follow these two cohorts into retirement in 2016 and 2018, respectively, and are used to report the employment choices made by these retirees. The two cohorts are combined for our analysis in the paper. This section describes the information from each cohort.

B.1.1 Survey of active workers in 2014 (S2014)

The survey population is comprised of workers that were aged 50 to 69, actively employed in March 2014, and have valid 2013 salary information indicating an active membership, according to administrative records as of March 2014. We sent a survey through email to those with an email address and by postal mail to the remaining workers. The final target population of 15,000 (9,012 and 5,988 surveys sent for the email and print sample, respectively). We received 2,622 responses (2,075 and 547 for email and print, respectively) for an overall response rate of about 18 percent.

B.1.2 Resurvey of workers (S2016-retiree)

The Retirement System provided us with administrative data on the full sample of 2014 active workers as of April 2016. We observe their benefit claiming status in the administrative records. The administrative records also include information on job classification, benefit type, benefit amount, and year of service. In May 2016, we sent a survey through email to all 2,622 respondents in S2014. The survey has two versions, actives and benefit claimants, which were sent to individuals with respective working status in the North Carolina Retirement System. Both active workers and benefit claimants are included in this paper. We received 1,208 responses with a response rate of 46.1%. Among them, 947 were still working as of 2016. Out of 947 workers who were working as of 2016, 76 responded to S2018 and self-identify as benefit claimants. We consider them part of the newly retired sample in this paper. The rest are included in the actives sample. We consider them part of the newly retired sample in this paper. 261 individuals started claiming benefits as of S2016, who belong to the newly retired sample.

B.1.3 Survey of active workers and benefit claimants in 2016 (S2016)

The survey population is comprised of workers that were actively employment in March 2014 according to administrative records as of March 2014 and were not deceased or claiming benefits other than TSERS/LGERS as of April 2016. We then exclude individuals who were in the target population of S2014. The final target population of 16,000 (11,170 active workers and 4,830 benefit claimants). We sent two versions of the survey, actives and benefit claimants, to individuals with respective working status in the North Carolina Retirement System according to the retirement system. Both active workers and benefit claimants are included in the paper. We received 4,200 responses for a response rate of about 26 percent. 3,311 were active workers as of April 2016, while 889 had started claiming benefits.

B.1.4 Resurvey of workers (S2018)

The Retirement System provided us with administrative data on the population of active workers and benefit claimants as of December 2017. The administrative records include information on job classification, benefit type, benefit amount, but not on benefit claiming date and year of service. We cannot observe the benefit claiming status as of May 2018. In May 2018, we sent a survey through email to all respondents with completed responses in S2016. The survey has two versions, actives and benefit claimants, and we asked respondents to self-identify as active workers or benefit claimants. Among the 5,408 workers in S2016 that we sent a survey to, 2,494 responded with a response rate of 46.1%. 1,445 of them were first surveyed in 2016 as active workers. Among them, 1,192 reported they were still working as of 2018, who belong to the active sample. The 253 individuals who self-classified themselves as benefit claimants are defined as part of the newly retired sample. 430 respondents of S2016 for benefit claimants stated that they had started claiming benefits as of 2018 and constitute the retiree sample.

Appendix Table B.1 shows how the three samples in our NC dataset are constructed. Check marks in the three columns under sample classification indicate how each subgroup is categorized.

B.2 Details of the Health and Retirement Study (HRS) dataset construction

The Health and Retirement Study (HRS) is a longitudinal survey on older individuals that collects detailed individual-level information including demographics, employment, pension, and health of respondents and their spouses, as well as financial characteristics such as income and wealth at the household level. The HRS is intended to provide data for research on the individual choices of people as they age and inform policy makers on the changing needs of the older population. Each HRS household observation must include at least one member satisfying the age eligibility requirements for each cohort. Starting in 1998, all cohorts are interviewed every two years. As of 2016, the initial HRS sample have been subject to 13 waves of surveys. The HRS is nationally representative study and uses a multi-stage national area probability sample of U.S. households with oversampling of Blacks, Hispanics and residents of the state of Florida. In additional to questionnaire sections that are included in each wave such as income and retirement, the HRS also has wave-specific experimental modules with additional questions that are administered to random subsamples.

In this study, we rely on the RAND HRS longitudinal file 2016 to construct our HRS dataset. It provides cleaned and streamlined information that are consistent through most HRS waves. The RAND file also imputes, when necessary, income, wealth, and medical expenditures variables based on the HRS Core Questionnaires.

Our primary unit of analysis is the household. Following the methodology in the RAND dataset, we create a household-level dataset by selecting all individuals in single-respondent households and just the Financial Respondent in couple households. Each household observation includes only one respondent and her spouse if married. Data construction thereafter focuses on the characteristics of the respondent in each household.

Our goal is to focus on individuals who transitioned into retirement during the years they were interviewed, with valid information pre- and post-retirement for us to analyze. Our definition of retirement is reporting "partly retired" or "retired" to the employment status questions. The dataset with all waves combined has 29,567 household-level observations. About 70% have reported being retired at least once. Among them, around 9% report having returned to full-time/part-time work in waves after first reporting retirement. Another 16% have reported employment status other than full-time/part-time work and retirement in waves after first reporting retirement. Therefore, we focus our analysis on survey results when individuals report being retired for the first time.

To construct our sample, we first determine the year when an individual first reports being retired according to their responses to the question on employment status. Among those who have ever reported being retired, we eliminate from our sample those without valid pre-retirement survey responses, including those who report retiring by the 1992 survey, the very first wave of the HRS. This gives us a sample size of 9,047. We also exclude those with missing information on pre- or post-retirement household income, bring the sample size down to 8,844. We also make the restriction that each observation in our sample has valid response in the 2010 wave. The final sample size is 6,518. Public sector workers are defined as having been employed by the federal government or a unit of a state, county, or local

government, or having worked in the public administration industry. The sample contains 2,061 public sector workers, which is around 32% of the sample.

Appendix Table B.3 present descriptive statistics for our sample, as well as all retirees in wave 2008, 2010, 2012, 2014, and 2016 respectively. Individuals in our sample are disproportionately younger and have higher post-retirement household income compared retirees in waves between 2008 and 2016. This is due to the fact that retirees in these waves on average retire much earlier than the retirement timing cutoff of our sample, 2008. For example, the average first-time retirement year is 1998 for all retirees in the 2008 wave. Other demographics are similar, although our sample has a slightly higher proportion of non-Hispanic African Americans and Hispanics.

Our key variable, household income, is constructed partly based on imputations in the RAND dataset. set to the sum of respondent and spouse earnings, pensions and annuities, SSI and Social Security Disability, Social Security retirement, unemployment and workers compensation, other government transfers, household capital income, and other income. Around 30% of observations in each wave have at least one income component imputed.

The wealth and income variables we analyzed were converted to constant 2016 dollars using the Personal Consumption Expenditures (PCE) price index issued by the Bureau of Economic Analysis (U.S. Department of Commerce), which is now favored by the Federal Reserve over the traditional adjustments based on the Consumer Price Index (CPI) released by the Bureau of Labor Statistics.

Information on the financial literacy of retirees is relatively limited in the HRS. The financial literacy questions are only available in experimental modules of the HRS in a few waves. Each respondent is randomly assigned to one and only one module. For example, the HRS 2010 wave has 10 experimental modules covering a variety of topics ranging from financial literacy to personality. 9.6% out of 22,034 respondents were randomly assigned to the financial literacy module. Apart from the 2010 wave, there are survey questions eliciting financial literacy in the "Retirement Planning" module of the 2004 wave, the "Financial sophistication and investment decision making" module of the 2008 wave, and the "Financial Mismanagement at Older Ages" module of the 2016 wave. The four waves all include the "Big Three" financial literacy questions developed by Lusardi and Mitchell (2011a) or questions with similar phrasing. After supplementing the 2010 wave financial literacy responses with the 2004, 2008, and 2016 wave, 22% of the HRS sample has non-missing information on financial literacy.

Starting from the 2008 wave, the HRS adds questions on credit card debt in the Asset, Debts section of the survey. The outcome variable on financial distress for the HRS sample is whether the respondent's household has carried over credit card balance from last month.

B.3 Key survey variables of the NC dataset

1. Household income question

   What is your household's total combined income for the past 12 months, approximately?

   Please include all wages, salary, commissions, bonuses, or tips from all jobs; net self-employment income from all businesses, proprietorships, and partnerships; interest, dividends, or net rental

income; income from pension benefits or Social Security benefits; income from any court-ordered alimony payments; and any other monetary income received by you and all members of your family who are 15 years of age or older. Include the annual benefit amount you received from either/both TSERS/LGERS pension retirement systems.

- Less than $25,000
- $25,000 to $49,999
- $50,000 to $74,999
- $75,000 to $99,999
- $100,000 to $149,999
- $150,000 to $249,999
- $250,000 or more
- Don't know

2. Supplemental Retirement Plan (SRP) balance question

   Which category best represents the total amount of all account balances of any retirement savings plans (e.g., 401(k), 457, 403(b), or IRA) that you currently hold? Do not include your TSERS/LGERS pension benefit here.

   - Less than $25,000
   - $25,000 to $99,999
   - $100,000 to $249,999
   - $250,000 to $499,999
   - $500,000 to $99,999
   - $1 million or more
   - No retirement savings plan
   - Don't know

3. Financial distress questions

   (a) Financial distress (credit card debt) (S2018 only)

   Over the past year, did you do any of the following when your spending exceeded your income? (Check all that apply.)

   - Spent out of savings/investments
   - Borrowed money from friends or family
   - Kept an account balance on a credit card
   - Got behind on payments or did not pay bills

- Other (please specify)
- My spending never exceeded my income

I define financial distress (credit card debt) as having chosen option "Kept an account balance on a credit card" in S2018.

(b) Financial distress (emergency funds)

How confident are you that you could come up with $2,000 if an unexpected need arose within the next month?

- I am certain I could come up with $2,000.
- I could probably come up with $2,000.
- I probably could not come up with $2,000.
- I am certain I could not come up with $2,000.

I define financial distress (emergency funds) as having chosen the last two options in each survey.

4. Financial literacy questions, "Big Three" questions developed by Lusardi and Mitchell (2011a)

(a) Financial Knowledge: Compounding

If you have $100 in your savings account, and the annual interest rate is 2%, how much money will you have in your account after five years?

- More than $110 *[Correct answer]*
- Exactly $110
- More than $102, but less than $110
- Exactly $102
- Less than $102
- Don't know

(b) Financial Knowledge: Inflation

Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?

- More than today
- Exactly the same
- Less than today *[Correct answer]*
- Do not know

(c) Financial Knowledge: Stock returns

Is this statement True or False? Buying a single company's stock usually provides a safer return than a stock mutual fund.

- True
- False *[Correct answer]*

Table B.1 Income changes and financial distress (HRS sample of public-sector workers)

Survey responses	N	Sample Classification: Transitioned to retirement between surveys "newly retired"	Actively working in surveys "working"	Retired in surveys "retiree"
S2014 responses, all working	2,622	-	-	-
Responded to S2016	1,208	-	-	-
Still working by S2016	947		✓	
		(except for 76 obs. who retired by S2018)		
Responded to S2018, still working	410		✓	
Responded to S2018, retired	76	✓		
Started claiming benefits by S2016	261	✓		
Responded to S2018, still retired	133	✓		
S2016 responses (new to surveys)	4,200	-	-	-
Working by S2016	3,311	-	-	-
Responded to S2018	1,445	-	-	-
Working by S2018	1,192		✓	
Started claiming benefits by S2018	253	✓		
Claiming benefits by S2016	889	-	-	-
Responded to S2018	430	-	-	-
Working by S2018	9	-	-	-
Started claiming benefits by S2018	421			✓
N	3,074	590	2,063	421
Sample size with non-missing income	2,710	473	1,865	370

Table B.2 Descriptive statistics of the NC samples and the HRS sample

Pre-retirement wave characteristics	NC newly retired	NC working	NC retiree	HRS public-sector workers	HRS private-sector workers
Men	34.7%	31.3%	35.7%	48.0%	47.1%
Married women	45.7%	45.1%	43.0%	21.7%	22.3%
Non-married women	19.7%	23.6%	21.4%	30.3%	30.5%
Race: Non-Hispanic black	10.1%	13.8%	11.6%	10.7%	10.1%
Race: Hispanic	1.9%	1.9%	0.3%	4.9%	8.1%
Race: White	86.3%	81.0%	85.9%	80.8%	78.6%
Race: Other	1.7%	3.3%	2.2%	3.6%	3.2%
Has bachelor's degree or above	75.3%	67.0%	67.9%	45.0%	19.5%
Age at survey	60.1	57.9	62.0	62.9	63.6
Household income $\leq$ \$25K	0.8%	1.3%	3.2%	25.0%	38.7%
\$25K - \$50K	12.5%	17.3%	25.9%	20.3%	23.0%
\$50K - \$75K	22.2%	25.8%	24.6%	16.1%	14.2%
$\geq \$75K$	64.5%	55.6%	46.2%	38.6%	24.2%
N	473	1,865	370	2,061	4,457

Note: The HRS sample include respondents who reported being retired for the first time in the HRS surveys between 2010 and 2016, responded to the wave of survey prior to first-time retirement, had non-missing household income information for both pre-retirement and post-retirement waves, and was the Financial Respondent of the household if married. The HRS sample results are weighted.

Table B.3 Descriptive statistics of the HRS sample, compared with the 2008-2016 wave retirees

Mean post-retirement characteristics	HRS sample	2008 wave	2010 wave	2012 wave	2014 wave	2016 wave
Female	52.6%	52.5%	52.8%	54.3%	54.7%	54.2%
Race						
Non-Hispanic black	10.3%	9.8%	8.9%	11.4%	11.4%	11.1%
Hispanic	7.0%	5.3%	5.4%	6.5%	6.9%	7.1%
White	79.3%	82.6%	83.3%	79.5%	78.6%	78.9%
Other	3.3%	2.2%	2.3%	2.7%	3.0%	2.8%
Has Bachelor's degree or above	28.1%	21.5%	23.4%	23.2%	24.4%	26.1%
Married	55.4%	44.6%	44.4%	42.5%	43.3%	44.3%
Public worker	33.6%	32.1%	32.5%	31.5%	31.8%	33.1%
Age at survey	63.4	73.2	74.3	72.8	73.1	73.5
Household income	83,730.3	66,740.5	52,345.7	49,388.1	54,703.2	59,927.7
$\leq$\$25k	34.1%	43.3%	41.6%	42.4%	39.5%	37.5%
\$25K - \$50K	22.1%	29.2%	30.2%	27.7%	27.1%	28.0%
\$50K - \$75K	14.8%	12.4%	12.6%	14.0%	13.9%	13.4%
$\geq \$75K$	29.0%	14.7%	15.2%	15.2%	19.0%	20.8%
Missing	-	0.4%	0.4%	0.7%	0.4%	0.4%
N	6,518	7,835	7,771	8,587	8,661	8,591
Proportion of retirees	-	56.7%	42.5%	52.5%	56.0%	47.9%

Note: The HRS sample include respondents who reported being retired for the first time in the HRS surveys between 2010 and 2016, responded to the wave of survey prior to first-time retirement, had non-missing household income information for both pre-retirement and post-retirement waves, and was the Financial Respondent of the household if married. The 2010-2016 waves sample include all retirees who were the Financial Respondent of the household if married. All sample results are weighted.

Table B.4 Income changes and household SRP balance changes

Income changes	Proportion in sample	Post-retirement household SRP balance changes (IRA and Keogh only for HRS retirees) SRP balance (category) increases	SRP balance (category) decreases	SRP balance (category) stays the same
Panel A: Men (NC newly retired)				
Income category increases	2.9%	0.0%	0.0%	100.0%
Income category decreases	50.0%	18.0%	20.0%	62.0%
Income category stays the same	47.1%	28.3%	17.4%	54.3%
Pension replacement ≤ median	48.2%	20.5%	17.9%	61.5%
Pension replacement > median	51.8%	22.0%	19.5%	58.5%
Panel B: Married women (NC newly retired)				
Income category increases	6.3%	12.5%	12.5%	75.0%
Income category decreases	53.1%	21.2%	30.3%	48.5%
Income category stays the same	40.6%	22.0%	26.0%	52.0%
Pension replacement ≤ median	39.1%	16.2%	24.3%	59.5%
Pension replacement > median	60.9%	19.7%	31.1%	49.2%
Panel C: Non-married women (NC newly retired)				
Income category increases	4.9%	0.0%	0.0%	100.0%
Income category decreases	54.1%	12.1%	33.3%	54.5%
Income category stays the same	41.0%	29.2%	20.8%	50.0%
Pension replacement ≤ median	42.9%	15.8%	42.1%	42.1%
Pension replacement > median	57.1%	7.1%	25.0%	67.9%
Panel D: NC working				
Income category increases	22.5%	38.9%	10.6%	50.5%
Income category decreases	10.4%	27.1%	24.8%	48.1%
Income category stays the same	67.1%	28.1%	13.6%	58.3%
Panel E: NC retirees				
Income category increases	24.1%	25.3%	16.5%	58.2%
Income category decreases	24.6%	20.5%	24.1%	55.4%
Income category stays the same	51.4%	18.9%	13.0%	68.0%
Pension replacement ≤ median	53.0%	23.4%	14.9%	61.7%
Pension replacement > median	47.0%	17.9%	18.6%	63.5%
Panel F: HRS first-time retiree (public sector)				
Income increases	36.9%	61.1%	38.9%	0.0%
Income decreases	63.1%	64.4%	35.4%	0.2%
Pension benefit is zero	61.2%	60.9%	38.9%	0.2%
Pension replacement ≤ median	13.4%	70.6%	29.4%	0.0%
Pension replacement > median	25.4%	64.7%	35.3%	0.0%

Note: This table does not contain NC workers who transitioned into retirement between S2014 and S2016, as S2014 does not contain information on SRP balance categories. HRS sample results are weighted.

Table B.5 Income change and financial literacy

	Number of correct answers to S2016 financial literacy questions					
Income changes	0	1	2	3	Don't know	Mean
Panel A: Men (NC newly retired)						
Income category increases	0.0%	0.0%	0.0%	100.0%	0.0%	3.0
Income category decreases	0.0%	2.0%	15.7%	78.4%	3.9%	2.7
Income category stays the same	0.0%	6.2%	33.3%	60.4%	0.0%	2.5
Pension replacement ≤ median	0.0%	7.5%	10.0%	80.0%	2.5%	2.7
Pension replacement > median	0.0%	0.0%	39.5%	60.5%	0.0%	2.6
Panel B: Married women (NC newly retired)						
Income category increases	0.0%	33.3%	33.3%	33.3%	0.0%	2.0
Income category decreases	2.7%	16.2%	35.1%	39.2%	6.8%	2.0
Income category stays the same	0.0%	21.1%	35.1%	42.1%	1.8%	2.2
Pension replacement ≤ median	2.3%	18.2%	31.8%	45.5%	2.3%	2.2
Pension replacement > median	0.0%	20.6%	36.8%	35.3%	7.4%	2.0
Panel C: Non-married women (NC newly retired)						
Income category increases	0.0%	0.0%	50.0%	50.0%	0.0%	2.5
Income category decreases	0.0%	9.1%	36.4%	54.5%	0.0%	2.5
Income category stays the same	4.0%	12.0%	32.0%	52.0%	0.0%	2.3
Pension replacement ≤ median	0.0%	0.0%	30.0%	70.0%	0.0%	2.7
Pension replacement > median	0.0%	14.3%	42.9%	42.9%	0.0%	2.3
Panel D: NC working						
Income category increases	1.8%	15.2%	35.8%	45.4%	1.8%	2.2
Income category decreases	3.4%	10.9%	29.9%	54.4%	1.4%	2.3
Income category stays the same	3.7%	16.1%	33.0%	44.8%	2.4%	2.2
Panel E: NC retirees						
Income category increases	2.4%	22.4%	30.6%	44.7%	0.0%	2.2
Income category decreases	3.4%	15.7%	30.3%	50.6%	0.0%	2.3
Income category stays the same	2.2%	14.1%	35.1%	48.6%	0.0%	2.3
Pension replacement ≤ median	2.6%	17.2%	31.8%	48.4%	0.0%	2.3
Pension replacement > median	2.4%	15.6%	34.1%	47.9%	0.0%	2.3
Panel F: HRS first-time retiree (public sector)						
Income increases	3.2%	13.8%	33.0%	49.9%	-	2.3
Income decreases	2.7%	14.4%	24.5%	58.4%	-	2.4
Pension benefit is zero	2.6%	16.1%	28.7%	52.6%	-	2.3
Pension replacement ≤ median	1.9%	6.9%	23.3%	67.9%	-	2.6
Pension replacement > median	4.0%	13.4%	25.8%	56.8%	-	2.3

Note: This table does not contain those who transitioned into retirement between S2014 and S2016, as S2014 contains only two questions on financial literacy. The sample contains workers with non-missing responses to financial literacy questions in both S2016 and S2018. The HRS sample only contain those with valid response for the financial literacy experimental modules. See Appendix B.3 for detailed information on financial literacy questions in the HRS. HRS sample results are weighted.

Table B.6 Income changes and financial literacy changes

Income changes	Proportion in sample	Change in financial literacy between waves: Number of correct answers increases	Number of correct answers decreases	Number of correct answers stays the same
Panel A: Men (NC newly retired)				
Income category increases	2.9%	0.0%	33.3%	66.7%
Income category decreases	50.0%	7.8%	19.6%	72.5%
Income category stays the same	47.1%	16.7%	12.5%	70.8%
Pension replacement ≤ median	48.2%	7.5%	20.0%	72.5%
Pension replacement > median	51.8%	18.6%	14.0%	67.4%
Panel B: Married women (NC newly retired)				
Income category increases	6.4%	11.1%	55.6%	33.3%
Income category decreases	52.9%	28.4%	24.3%	47.3%
Income category stays the same	40.7%	24.6%	14.0%	61.4%
Pension replacement ≤ median	39.3%	20.5%	27.3%	52.3%
Pension replacement > median	60.7%	32.4%	22.1%	45.6%
Panel C: Non-married women (NC newly retired)				
Income category increases	3.3%	0.0%	0.0%	100.0%
Income category decreases	55.0%	6.1%	33.3%	60.6%
Income category stays the same	41.7%	24.0%	20.0%	56.0%
Pension replacement ≤ median	41.7%	5.0%	25.0%	70.0%
Pension replacement > median	58.3%	21.4%	32.1%	46.4%
Panel D: NC working				
Income category increases	22.1%	22.7%	23.3%	54.0%
Income category decreases	9.7%	19.7%	33.3%	46.9%
Income category stays the same	68.1%	22.4%	22.2%	55.4%
Panel E: NC retirees				
Income category increases	23.7%	17.6%	21.2%	61.2%
Income category decreases	24.8%	14.6%	36.0%	49.4%
Income category stays the same	51.5%	16.8%	23.2%	60.0%
Pension replacement ≤ median	53.5%	18.2%	23.4%	58.3%
Pension replacement > median	46.5%	14.4%	28.7%	56.9%

Note: This table does not contain those who transitioned into retirement between S2014 and S2016, as S2014 contains only two questions on financial literacy. The sample contains workers with non-missing responses to financial literacy questions in both S2016 and S2018.

Table B.7 Regression of income changes - pension benefit replacement rate (NC sample)

Respondent / household	Pension benefit replacement is above median			
characteristics in S2014/S2016	(1)	(2)	(3)	(4)
Married female	0.133***	0.136***	0.119**	0.121**
	(0.045)	(0.046)	(0.048)	(0.048)
Non-married female	0.093*	0.083	0.070	0.070
	(0.054)	(0.055)	(0.055)	(0.062)
African American	0.073	0.080	0.072	0.071
	(0.064)	(0.065)	(0.066)	(0.066)
Hispanic	-0.189	-0.197	-0.215	-0.214
	(0.156)	(0.157)	(0.154)	(0.154)
Bachelor's degree or above	-0.045	-0.055	-0.036	-0.039
	(0.049)	(0.052)	(0.053)	(0.053)
Age at survey	-0.048***	-0.050***	-0.050***	-0.049***
	(0.005)	(0.005)	(0.005)	(0.005)
Has good health	-0.021	-0.055	-0.072	-0.069
	(0.051)	(0.053)	(0.055)	(0.055)
Transitioned to retirement	0.100**	0.050	0.052	0.049
between S2016 and S2018	(0.043)	(0.089)	(0.089)	(0.089)
Salary in 2nd quartile	0.169**	0.192***	0.190***	0.187**
	(0.069)	(0.070)	(0.070)	(0.070)
Salary in 3rd quartile	0.330***	0.354***	0.358***	0.359***
	(0.063)	(0.063)	(0.063)	(0.063)
Salary in 4th quartile	0.378***	0.406***	0.413***	0.415***
	(0.062)	(0.062)	(0.062)	(0.062)
Financial literacy questions all correct			-0.042	-0.042
			(0.043)	(0.043)
Financial literacy answers missing			0.174	0.176
			(0.141)	(0.141)
Self-rated financial knowledge is high			-0.061	-0.062
			(0.043)	(0.043)
Has supplemental retirement saving accounts			-0.005	-0.005
			(0.060)	(0.060)
Spouse currently employed				0.025
				(0.046)
Spouse covered by pension plan				-0.034
				(0.046)
Control for agency type and retirement year	No	Yes	Yes	Yes
Sample size	777	777	777	777
Mean dependent variable	0.498	0.498	0.498	0.498

Note: The sample contains retired workers with non-missing records for net benefit paid out per month. Average marginal effects from probit models are reported. Salary is the 2013 annual salary based on administrative records. Financial literacy information in S2016 is used in models. * p<.10, ** p<.05, *** p<.01.

Table B.8 Regression of financial distress (credit card debt) on replacement rate (NC sample)

Respondent / household	Kept credit card balance when spending exceeded income			
characteristics in S2014/S2016	(1)	(2)	(3)	(4)
Pension replacement rate above	0.011	-0.003	-0.004	-0.003
median	(0.040)	(0.040)	(0.040)	(0.041)
Married female	0.062	0.068	0.057	0.074
	(0.053)	(0.053)	(0.054)	(0.056)
Non-married female	0.100**	0.084*	0.073	0.071
	(0.048)	(0.048)	(0.049)	(0.049)
African American	-0.037	-0.050	-0.050	-0.053
	(0.057)	(0.054)	(0.055)	(0.054)
Hispanic	0.226	0.235	0.242	0.233
	(0.176)	(0.179)	(0.180)	(0.183)
Age at survey	-0.004	-0.003	-0.002	0.000
	(0.005)	(0.005)	(0.005)	(0.006)
Transitioned to retirement	-0.028	-0.022	-0.018	-0.022
between S2016 and S2018	(0.079)	(0.079)	(0.079)	(0.080)
Household income $50-75k	0.028	0.042	0.046	0.053
	(0.048)	(0.049)	(0.049)	(0.051)
Household income $\geq \$75k$	-0.056	-0.004	0.005	-0.013
	(0.046)	(0.048)	(0.049)	(0.054)
Has supplemental retirement saving		-0.099*	-0.086	-0.085
accounts		(0.058)	(0.058)	(0.058)
SRP balance $\geq$ $100k		-0.146***	-0.143***	-0.137***
		(0.039)	(0.039)	(0.040)
Financial literacy questions all			-0.017	-0.016
correct			(0.038)	(0.038)
Self-rated financial knowledge is			-0.084**	-0.077*
high			(0.038)	(0.039)
Spouse currently employed				0.021
				(0.045)
Spouse covered by pension plan				0.042
				(0.044)
Spouse started claiming SS benefits				-0.107**
				(0.041)
Working after retirement				-0.006
				(0.040)
Has started claiming SS benefits				0.014
				(0.051)

Note: The sample contains retired workers with non-missing records for net benefit paid out per month and a valid response to S2018, as financial distress measured by keeping credit card balance is only available in S2018. Sample size is 600. Average marginal effects from probit models are reported. Mean dependent variable is 0.254. Financial literacy information in S2016 is used in models. Specifications also include bachelor's degree, years since retirement, self-rated health, and agency controls. * p<.10, ** p<.05, *** p<.01.