Determining School-Level Proficiencies

During the time period of this study, the percentage of students needing to score proficient to achieve AYP increased from 71% to 77% in language arts and 64% to 71% in mathematics. The determinations of whether schools met these AYP benchmarks for proficiency were based on the whole school population and did not include calculations for each subgroup required under NCLB. Furthermore, the increase in proficiency benchmarks (from 71% to 77% in language arts and from 64% to 71% in mathematics) during the 2006-07 school year did not affect any elementary school’s AYP status based on whole school proficiency calculations. In other words, no schools in the district met the lower benchmark but failed to meet the higher benchmark when it was increased during the 2006-07 school year. This also indicates that our “lower performing” and “higher performing” school measures are stable during the years of our study, and no schools switch from higher performing status to lower performing status (or vice versa) during the timeframe of our study.

Variable Coding

In order to accurately predict whether or not intra-district transfer increases student achievement, we used several student background measures to account for differences between choosers and non-choosers and also to account for general differences between students that might also influence achievement. These measures included gender, SES, race/ethnicity, English language proficiency, disability status, guardianship, neighborhood, number of years in the district, grade level, and prior achievement.

Gender was a dichotomous variable coded 1 for female and 0 for male. Free and reduced lunch status was used as a proxy for family SES, with free and reduced lunch participants coded 1 and non participants coded 0. Race/ethnicity was measured with a series of dummy variables: Asian, African-American/Black, Native American, Hispanic/Latino, Pacific Islander, and white (reference group). Language proficiency was coded 1 for students who were considered English language learners and 0 for students who were proficient in English. Students with severe disabilities were excluded from analyses because they qualify for specialized choice programs that are not considered intra-district transfers; however, we measured disability status for students with mild and moderate disabilities such as dyslexia or attention deficit disorder with a dichotomous variable coded 1 if the students had a documented disability and 0 if they did not. Students’ guardianship was also included as a series of dummy variables: the student lives in a two-parent home (reference group), the student lived in a single-parent home, and the student was cared for by someone other than their parents. The number of years each student attended a school in the district was also included as a continuous measure, and student grade level was calculated using a series of dummy variables for fourth, fifth, and sixth grade. Fourth grade was used as the reference category.
Because students were zoned to schools based on where they live and because residential areas within this district were segregated by race/ethnicity and also by social class, we included measures of students’ neighborhoods to account for possible neighborhood effects (Crane, 1991; Jencks & Meyer, 1990; Levine & Levine, 1996). We did so by creating a series of dummy variables indicating membership in one of 27 elementary school zones. Because students in the district were zoned to neighborhood schools (i.e., the schools that were closest to their homes), we used the boundaries of the school zones as indicators of neighborhoods. A small number of students (about 1-2%, depending on the year) were either homeless or did not reside within the district boundaries. As such, they were not assigned an attendance zone by the district. Because their school zone information was missing from our data, these students were excluded from analyses.

We also included a measure of prior achievement as a control variable in our analyses. Prior achievement provided us with a test score that reflects the history of all those factors affecting the student’s cumulative, retained learning up to that point in time, including early childhood events, the history of family and other environmental factors, the historical flow of school inputs, and any other factor that was both related to student achievement and did not change between the time points in which achievement was measured (Boyd et al., 2008).

Because state tests were administered in the spring of each academic year, we used the previous year’s spring achievement scores as our measure of prior achievement in both language arts and mathematics. As with our outcome measures of achievement, prior achievement was standardized and converted into z-scores by subtracting the mean from each score and dividing the product by the standard deviation. Means and standard deviations were computed separately by grade, year, and subject (language arts and mathematics). Due to the distributional properties of these scores for students in first and second grade, all students enrolled in grades two and under were not considered accurate measures of prior achievement and were therefore excluded from analyses. Students with missing test scores were also omitted from analyses.

**Propensity Score Matching**

To account for the effects of non-random assignment in the school choice process, researchers have utilized a variety of methodological and statistical techniques such as randomized field trials (Greene, Peterson, & Du, 1997; Hoxby & Rockoff, 2004; Kemple, 2001; Wolf et al., 2009), student-level fixed effects modeling (Bifulco & Ladd, 2006; Eberts & Hollenbeck, 2002; Hanushek, Kain, & Rivkin, 2002; Sass, 2006; Solomon, Paark, & Garcia, 2001), and instrumental variables (Cullen, Jacob, & Levitt, 2005; Ozek, 2009; Rouse, 1998). However, the use of each of these methods requires specific types of data or research contexts that are unavailable in our study, primarily because everyone who exercised choice in this particular district received their first choice option and because most students exercised choice in the first or second year of their elementary schooling experience. For example, randomized field trials require a control group generally created from a random selection of students who did not receive admittance into their school of choice, a condition which is unavailable in our study. When studying school choice, student-level fixed effects models imply that students exercise choice at some time after they have at least a baseline measure of achievement in a non-choice school. However, in our study, the majority of elementary choice participants chose schools prior to the years in which we have accurate test score data. In this district, elementary intra-district transfer participants generally exercise choice during the first few years of enrollment. About 50% of choosers exercise choice in kindergarten. Another 30% choose when they enter first grade. The remaining 20% choose during second through sixth grade. Relatively few students exercise choice during third through sixth grades. Again, in a study where no lottery was employed when determining which students were able to attend their school of choice, instrumental variables are difficult to create. Because our study is unique in that every student who exercised choice
was able to enroll in his or her first-choice school, and because most intra-district transfer participants exercised choice at periods of time that coincide with natural school changes (e.g., when students enter school for the first time), we were unable to use the most common methodological and statistical techniques that are often employed to eliminate the association between the treatment variable of interest and the error term.

To account for the non-random assignment of students to the “treatment” used in this study (intra-district transfer participation), we use propensity score matching (PSM). PSM provides a mechanism in which one can adjust for selection bias by summarizing covariates about treatments in a graduated arrangement, which allows for causal inference when comparing treatments to controls (Rosenbaum, 1995). Morgan and Harding (2006) also argue that all matching estimators can be thought of as reweighting schemes whereby treatment and non-treatment observations are reweighted to all causal inferences on the difference in means. As such, the advantages of propensity score matching include the minimization of differences on all covariates, which addresses selection issues regarding the relationship between a treatment and an outcome (Rosenbaum, 1995).

Ideally, a propensity score would incorporate all information about selection; however, practically, the propensity score must be estimated using only available and observed variables. As a result, the PSM estimator can be biased and inefficient (Morgan & Harding, 2006; Morgan & Winship, 2007; Rosenbaum & Rubin, 1983; Rosenbaum, 1995). To minimize bias without increasing variance, we employed matching techniques and algorithms as they are suggested in the literature. Using experimental data with which matching estimators can be compared, researchers argue for the advantages and efficiencies of kernel matching (Heckman, Ichimura, Smith, & Todd, 1998; Heckman, Ichimura, & Todd, 1997; 1998; Smith & Todd, 2005). Specifically, we use the Epanechnikov kernel matching algorithm, in which all members of the non-treated group are used to build a match for each member of the treated group. In other words, the kernel is a function that weights the contribution of each non-treated group member according to the distance between their propensity scores and the propensity scores of members of the treatment group. Exact matches receive larger weights, and poor matches receive smaller weights. Using this algorithm is ideal for data such as ours where the treatment group is smaller than the non-treated group, and many of the non-treated cases could potentially provide good matches for members of the treatment group. Using and reweighting all available cases also decreases the possibility of discarding important information when using other algorithms that only identify one match per treated case, which some researchers suggest is an improvement over one-to-one matching schemes (see Rosenbaum, 1995).

To create our match, we used Leuven and Sainesi’s (2003) “psmatch2” module for Stata 11 statistical software to estimate propensity scores for both treatments (i.e., intra-district transfer participants) and non-treatments (i.e., non-participants) using a logit model. To avoid the potential for poor matches across grade, year, and subject area, we ran separate propensity score models for each grade, within each year, within each subject area (i.e., language arts and mathematics). The propensity score specification included measures of gender, SES, race/ethnicity, English language proficiency, disability status, guardianship, neighborhood, number of years in the district, and prior achievement. In this study our measure of prior achievement does not represent a true counterfactual because in most cases, it was not measured prior to students’ encounters with their choice schools. As a result, it could potentially introduce bias into our results. To address this issue, we computed both our propensity scores and our analyses with and without prior achievement as a covariate. In our study, the results associated with the effects of intra-district transfers on achievement were similar both with and without accounting for prior achievement. Due to these similarities we chose to include prior achievement in our analyses. These covariates are all reported as having some relationship between school choice and student achievement and therefore could potentially confound the effect of intra-district transfer on achievement. As Rubin and Thomas (1996) and Morgan (2001) have addressed, an important factor in determining whether or not to include variables and
higher-order terms in the propensity score model is not their statistical significance, but their power in balancing the means and variances of the covariates between treated and untreated groups.

In this study we used a common-support match, meaning that treated students who did not have a match within a specified interval on the propensity score were not paired with a student from the non-treated group. These cases were considered “off support” and were not included in our analyses. In the language arts combined sample (years 2004-06 to 2006-07 combined for grades four through six), 28 treated cases were “off support”, and 19 treated cases in the combined mathematics sample were “off support”. The distribution of propensity scores in the language arts sample ranges from .00 to .87 with a mean of .15. Similarly, the distribution of propensity scores for the mathematics sample ranges from .00 to .89 with a mean of .15. Additional information outlining the quality of the match, including estimates of bias reduction produced through the matching process and individual sample t-test results comparing the data before and after the match was performed all demonstrated a successful match. After deleting all cases that were “off support”, the language arts sample included 10,372 cases, with 1,534 treated and 8,866 untreated cases. The mathematics sample included 10,337 cases, with 1,536 treated and 8,820 untreated.

**Modeling & Weighting Procedures**

To determined the effects of intra-district transfer on student achievement, we ran two separate sets of linear regression models: one with language arts achievement as the outcome and another with mathematics achievement as the outcome measure of interest. We also included all covariates in our final models which used intra-district transfer participation to predict student achievement in both language arts and mathematics. These covariates included prior achievement, student gender, SES, race, language learning status, disability status, guardianship, grade level, and number of year the student attended a school in the district. In addition to these student background variables, we also included a measure of time to measure achievement growth as well as an interaction between time and intra-district transfer participation to assess the degree to which the effects intra-district transfers on student achievement grow over time.

To minimize the differences on all covariates that were related to both intra-district transfer and student achievement, each model was weighted using the sampling weights generated by the PSM procedures. Weighting the analyses by the sampling weights generated as part of the PSM procedure should eliminate the need to control for the covariates used to create the initial propensity score match. Even though these controls are related to the student achievement outcomes used in this study, their presence in the models should no longer influence the effect size of our treatment variable (intra-district transfer participation). To test this, we ran a series of step-wise regression models, introducing control variables in blocks. Through this process we determined that our match was successful in eliminating endogeneity in our models and that the full model that included all covariates was similar to the simple model which simply used intra-district transfers to predict achievement. In an effort to maximize the information presented in this chapter, we report the results from our full models.