Evidence of Learning Gains at Seattle's South Shore School

November 2013

Prepared for:

The League of Education Voters

ECONorthwest
ECONOMICS • FINANCE • PLANNING

Contact Information

ECONorthwest specializes in economics, planning, and finance. Established in 1974, ECONorthwest has over three decades of experience helping clients make sound decisions based on rigorous economic, planning, and financial analysis.

Bryce Ward and Melissa Rowe prepared this report. ECONorthwest is solely responsible for its content.

ECONorthwest gratefully acknowledges the substantial assistance provided by staff at Seattle Public Schools and the Washington State Education Research and Data Center.

For more information about ECONorthwest, visit our website at www.econw.com.

For more information about this report, please contact:

ECONorthwest
222 SW Columbia Street
Portland, OR 97201
503-222-6060
econ@econw.com

Table of Contents

Summary

The South Shore School, a public PreK-8 school in southeast Seattle, is the result of an innovative public/private partnership between the Seattle Public Schools district and the New School Foundation (now merged with the League of Education Voters Foundation). South Shore's program is based on rigorous, evidence-based interventions, including high-quality and integrated pre-Kindergarten, small class sizes in early grades, and extra supports for students and faculty.

ECONorthwest conducted an evaluation of South Shore's impacts on student performance, comparing South Shore student performance to that of observably similar Seattle Public Schools students.

Examining student-level test data across multiple school years, ECONorthwest finds that attendance at South Shore is generally associated with improved scores on the WASL (Washington Assessment of Student Learning) and MSP (Measurements of Student Progress) assessments for grades 3-8 and an increased likelihood of meeting the state standards in math and reading. The South Shore effects are large and statistically meaningful, particularly in math.

The effects of attending South Shore are larger for students who enrolled during pre-Kindergarten or Kindergarten, and effects on achievement are consistent across racial and ethnic groups.

What is the South Shore School?

The South Shore School, a public PreK-8 school in southeast Seattle, is the result of an innovative public/private partnership between the Seattle Public Schools district and the New School Foundation that began in 1998. South Shore's program is based on rigorous, evidence-based interventions, including high-quality and integrated pre-Kindergarten, small class sizes in early grades, and extra supports for students and faculty.[1]

In 2011, the New School Foundation merged with the League of Education Voters (LEV) Foundation, which now administers an annual grant to South Shore and works statewide on policy, advocacy, and implementation related to PreK-3rd alignment projects.

In the 2011-12 school year, South Shore enrolled approximately 700 students in grades pre-K through 8. Students at South Shore are predominantly African-American and Asian and are drawn mostly from southeastern Seattle.

Purpose of this Report

LEV commissioned ECONorthwest to provide an updated evaluation of the effectiveness of South Shore's program. ECONorthwest completed similar evaluations twice before, in 2009 and 2010.

In this report, as in our previous analyses, we present the results of an analysis that is more rigorous than simple cross-school comparisons, although it stops well short of the "gold standard" of randomized controlled trials. Specifically, we use student-level demographic and educational data provided by the Seattle Public Schools district (SPS) to compare South Shore students to similar students attending other SPS schools.

Our analysis was designed to answer the following questions:

- Do South Shore students perform better than observably similar SPS students?
- Are there meaningful differences in student performance between long-term students and new arrivals?

Methodology and Data Sources

Fundamentally, an evaluation of South Shore (or any school) consists of comparing the outcomes of South Shore students to the outcomes of students that did not attend South Shore. However, it is not particularly informative to simply compare the outcomes of all South Shore students to all other SPS students or the students attending other specific schools. Raw differences between South Shore outcomes and the outcomes at other schools reflect much more than differences in the educational environment. In addition to school effectiveness, student performance depends on individual, family, and other contextual characteristics. As such, it is possible (even likely) that two equally effective schools will have significantly different student outcomes due to large differences in the composition of their student populations.

The goal of this evaluation is to isolate the effect of South Shore by accounting for underlying differences that affect academic performance, such as student demographic characteristics, family situations, and other contextual factors.

Researchers face two challenges when undertaking such evaluations. First, school performance data are noisy (i.e., they can fluctuate fairly widely from year to year for reasons unrelated to school quality).[2] Such fluctuations can reflect a variety of factors, such as random differences among cohorts, changes to the composition of the cohort as students move in and out, or a dog barking outside while students are completing an assessment. These random fluctuations can make it difficult to accurately measure school performance. To minimize the influence of random fluctuations, we examine multiple years of data in each analysis described in this report.

Second, no other school (or collection of schools) provides a perfect comparison for a given school and its programs. That is, there is no school that is identical to South Shore in every way except for its South Shore programs. Student populations differ across schools, and the available data do not capture all relevant differences between students. For example, students with more-educated parents tend to perform better in school. As such, if the data analyzed do not include and account for parents' education, an analysis might falsely attribute certain outcomes stemming from differences in parents to differences in school quality.

The ideal way to accurately measure school performance is through a randomized controlled trial. In the case of South Shore, using this methodology would involve randomly assigning a large population of students to South Shore (the treatment group) or to other SPS schools (the control group). Under this scenario, there would ideally be no underlying, unobserved differences between the treatment and control groups (i.e., membership in each group would be random).[3] As such, any differences between the two groups could be clearly attributed to attendance at South Shore.

In practice, students choose (or their parents choose for them) to attend South Shore, either explicitly by selecting South Shore in SPS's school choice program, or implicitly by choosing to live within the South Shore's neighborhood-school boundaries. Because of this "self-selection" into South Shore, we cannot entirely rule out the possibility that, compared with SPS students overall, South Shore students have underlying differences that affect their performance in school. However, we can minimize the possibility that our results reflect such underlying differences by obtaining additional information about South Shore students and SPS students.

Specifically, we can use statistical information about characteristics that are typically correlated with performance on achievement tests, such as race, socioeconomic status, and attendance, to statistically control for the effects of these underlying differences using regression analysis.[4] This method allows us to compare groups of students who have mostly similar observable, relevant, and measurable characteristics except for their exposure to "treatment," in this case, South Shore attendance.

But even after accounting for observable differences, differences between South Shore students and non-South Shore students could remain. While our analytical methodology is subject to more error than a randomized controlled trial methodology, it nevertheless helps us understand the relationships between relevant characteristics that we can observe and measure. Appendix B includes additional detail about comparison populations and effect measurement.

Compared to What? Two Approaches to Examining Student Performance

Even after accounting for differences among students, evaluating school performance requires some criteria to determine what constitutes good performance. That is, student and school outcomes need to be placed in a larger context that helps people understand what the results mean. Typically, there are two ways to judge performance:

- **Absolutely**—This approach compares student performance to a set of predetermined benchmarks
- **Relatively**—This approach compares student performance to the performance of other students in the same grade and year

Each of these approaches has important strengths and weaknesses. If the benchmarks are clearly understood, the absolute approach provides an easily interpretable metric for school performance—X% of students met the defined standards. This simple comparison allows performance to be easily compared across time and space (assuming testing or other measurement tools—and the benchmarks—remain constant).

The problem with the absolute approach is that it ignores what is happening away from the benchmark. Students can exceed or fall short of the benchmark by wide margins. The absolute approach ignores this information (and thus creates incentives for schools to divert resources from those students who clearly fall above or below the line to those who are likely to fall near the line). It is possible for two schools to have identical shares of students meeting the benchmark, but drastically different overall performance. For instance, one school could have most of its students exceeding the benchmark by a wide margin, while the other could have most of its students just barely meeting the benchmark.

The relative approach helps address this problem. In the relative approach, the full distribution of student performance is used to evaluate student and school performance. Schools are evaluated based on their students' average performance relative to the average performance of students in the whole population. In this type of metric, every student's performance matters (not just those near the benchmark). Furthermore, because students are typically compared relative to those in the same year and in the same grade, this type of metric does not require absolute consistency in the quality of the testing (and associated benchmarks).

The problem with the relative approach is that it is "zero-sum," so it may be difficult to interpret. When students and schools are moving relative to each other, for one student or school to move up, another must move down. As such, the relative approach

does not identify any widespread gains (or losses) in student performance. Thus, unless the distribution of student performance remains constant over time, it can be difficult to understand the information provided by this approach. A student with the same level of understanding and performance can receive very different relative scores from year to year.

Neither of these approaches is sufficient by itself to provide a full picture of student and school performance, so throughout this report we examine both how South Shore students perform relative to statewide benchmarks and how they perform relative to other SPS students in the same grade and year.

Acquiring and Organizing the Data

ECONorthwest's first task in conducting an evaluation of Seattle's South Shore School was to acquire, assemble, and organize the necessary data. This section describes that process generally; Appendix A provides additional detail about how we created the databases for our study.

Fundamentally, our analysis consists of examining the outcomes (e.g., test scores) for South Shore students and comparing them to similar students who did not attend South Shore. To complete such an analysis, we needed to obtain de-identified, student-level longitudinal data for South Shore students as well as SPS students for the same time period. SPS provided these data in the following categories: demographics, historical enrollment, coursework, state assessments, attendance, and school choice data.

We requested and received a similar dataset from the Washington State Education Research and Data Center (ERDC). The ERDC data contains most of the same elements as the SPS data, with the addition of FRL status and the exception of school choice data.

Assessment data

In this evaluation, we focus on student performance on state assessments. The Washington Assessment of Student Learning (WASL) was Washington's state test from 1997 until summer 2009. Students in grades 3-8 and 10 took the WASL each spring in reading and math. Students were also tested in writing in grades 4, 7, and 10, and science in grades 5, 8, and 10. Our previous reports considered only reading and math WASL scores because relatively few South Shore students had reached 4th and 5th grades.

In spring 2010, Washington State replaced the WASL with the Measurements of Student Progress (MSP) and the High School Proficiency Exams (HSPE). The MSP is for students in grades 3-8 for reading and math, and grades 4 and 7 for writing, and grades 5 and 8 for science. The HSPE measure the basic proficiency of high school students in reading and writing, and serve as the state's exit exams in those subjects.[5]

This analysis primarily uses WASL and MSP results, as only a small number of South Shore students (i.e., the 2002-03 Kindergarten cohort) were in 9th grade in 2011-12, the latest year for which we have data.

As discussed above, we examine performance both absolutely and relatively. To measure absolute performance, we

examine whether or not students met the state benchmark for reading or math in their grade. To measure relative performance, we calculate a standardized score—in standardized units (SUs)—for each student's test results. A standardized score reports how far, in terms of standard deviations on a normal curve, the student is from the average student in his or her grade and year. For the purposes of the charts and tables in this study, we analyze and display standardized WASL and MSP scores together.

Standardized scores (SU) are interpreted in a variety of ways based on contextual factors. For student-level interventions, What Works Clearinghouse considers effect sizes of at least 0.25 SU to be substantively important.[6]

Demographics, attendance, and school choice

To account for potential differences between students attending South Shore and those attending other SPS schools, we used the following information:

- Sex
- Ethnicity (as coded in seven categories: Asian, Black, Hispanic, Multi-ethnic, Native American, Pacific Islander, and White)
- Grade
- An indicator equal to one if a student is identified as residing in a non-English speaking household
- An indicator equal to one if the student is classified as living with both parents
- An indicator equal to one if the student has an IEP classification
- An indicator equal to one if the student is classified as gifted
- The number of SPS elementary or middle schools attended (a proxy for student mobility)
- School year
- School choice ranking: Seattle students are allowed to rank their preferences for the school they attend. Actual school assignment is largely a function of geography and sibling attendance. In some of our analyses, we limit the sample of students to include only those attending South Shore or who ranked it as their first choice at some point.

Data Limitations

The data used in this analysis have potential limitations that should be taken into consideration in interpreting the results. First, if the WASL or MSP tests do not accurately measure student ability, the results of the analysis will be affected. Second, the small sample sizes inherent in the South Shore population increase the possibility of measurement and sampling error for key variables.

Two other important analytical issues must be considered. First, as discussed previously, this analysis is subject to selection bias because students are not randomly assigned to either South Shore or any other school. The factors that cause a student to apply to and attend South Shore may be a result of unobserved underlying differences in those students compared to those who do not apply. Therefore, the analytical results may reflect unobserved differences in student characteristics rather than, or in addition to, the observed differences that are accounted for in the regression analysis.

Second, the analysis does not measure the effects of specific South Shore programs. We simply calculate the total effects of South Shore on student achievement, accounting for the observable factors outlined above. These effects are what parents are most likely to be concerned with. That is, they reflect the expected change in student performance contingent on attending the school. We do not attempt to explain precisely which policies produce which effects, although this is what administrators are most likely to care about. We also cannot rule out the possibility that the results may reflect differences in context (or peer effects) and not policy.

Findings

Overall, we find that South Shore students perform at levels significantly higher than would be predicted based solely on their observable characteristics.

In our previous studies, we reported that, relative to similar students attending other schools, South Shore 3rd and 4th grade students were more likely to meet the WASL benchmarks in both reading and math, and they had higher standardized test scores. The sizes of the effects were larger for students who enrolled during pre-Kindergarten or Kindergarten compared to those who enrolled later. And South Shore's effects were consistent across ethnic groups.

The current analysis finds similar results. Effects vary by grade, but South Shore students have higher test scores than demographically similar SPS students in math, reading, and writing. South Shore effects are particularly strong in math.

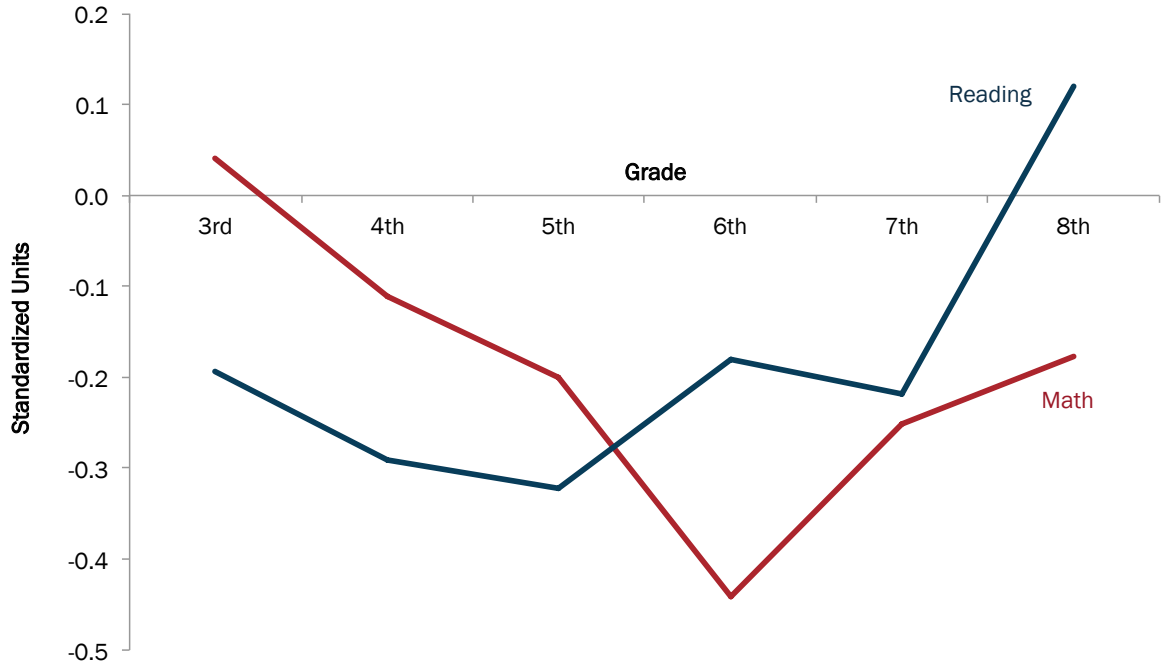
These effects are generally consistent across subgroups. We also find that students who attend South Shore in pre-Kindergarten and Kindergarten perform better than students who enroll in later grades and better than their demographically similar SPS counterparts, particularly in math.

We describe a simple set of results in the main body of the report. Appendix B contains additional analyses that use alternative estimation strategies and assumptions. The pattern of results described in Appendix B follows the same pattern as the results described here; however, the precise magnitude of the effects does vary somewhat.

Unadjusted Performance

Looking at overall performance on math and reading assessments without accounting for underlying differences between groups, the average South Shore student scores below the average SPS student in most grades (see Figure 1). In general, the average South Shore student performs slightly better than the average SPS student in 3rd grade math and 8th grade reading, and below the SPS average in every other grade.

Figure 1: Deviation of South Shore math and reading scores from Seattle Public Schools average



Source: ECONorthwest analysis of Seattle Public Schools data.

Observable Underlying Differences

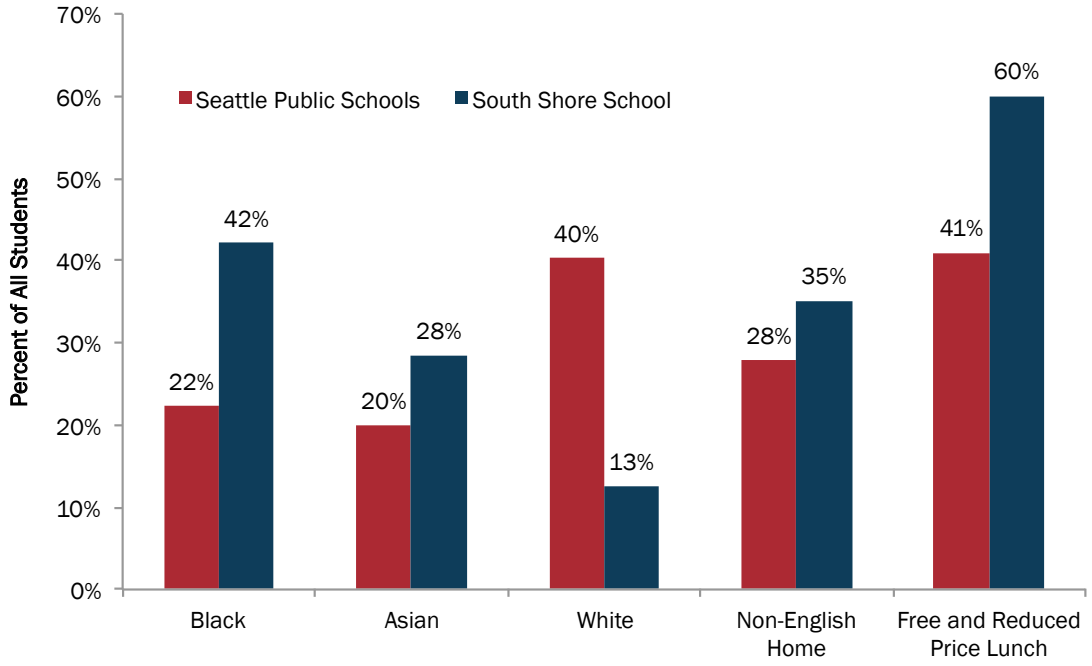
When comparing South Shore students with students at other schools, it is important to understand the observable underlying characteristics of the groups, because these may explain some of the differences in test scores. It is likely that at least some of the observed differences in school performance reflect differences in sociodemographic characteristics of South Shore students.

Figure 2 shows comparisons of demographic characteristics for South Shore students compared to SPS students overall.

Compared to SPS overall, South Shore has a larger share of black students and Asian students and a smaller share of white students and Hispanic students. While the SPS population is about 40 percent white, the South Shore population is about 13 percent white.

South Shore also has a slightly larger proportion of students living in non-English-speaking homes, and a significantly larger share of students receiving free or reduced price lunch.

Figure 2: Share of Seattle Public Schools students and South Shore School students with selected characteristics



Source: ECONorthwest analysis of Seattle Public Schools data and ERDC data.

Comparison of Outcomes Accounting for Observable Differences

In this section we examine differences in performance between South Shore students and students at all SPS schools after considering the underlying observable differences discussed above. To account for these potential differences, we estimate the effect of attending South Shore by estimating regression equations that include a variety of controls for student characteristics.[7] By doing so, we can assert with greater confidence that the differences between groups reported here are due to South Shore treatment effects and not to other observable differences.

3rd-8th grade assessment performance

Examining student-level test data across multiple school years, ECONorthwest finds that attendance at South Shore is associated with improved scores on WASL and MSP assessments in several grade and subject combinations, as well as an increased likelihood of meeting the state benchmarks in math and reading. These effects are large and statistically meaningful, particularly in math. Specific findings are discussed below.

Focusing on 3rd grade students, the first grade for which we have assessment scores and for which we have seven cohorts of students (about 375 students in total), we find that students enrolled at South Shore for 3rd grade score significantly higher in math (0.30 SU) and moderately higher in

reading (0.11 SU) than demographically comparable students attending other SPS schools (see Table 1 and Figures 3 and 4).

South Shore students are also significantly more likely to meet or exceed WASL and MSP benchmarks in both reading and math. Table 1 shows the share of South Shore 3rd graders we would expect to meet or exceed the benchmarks in math and reading based solely on student demographic characteristics, compared to the actual share of students meeting the benchmarks. The table indicates that South Shore students scored 12 percentage points above their expected scores in math and 5 percentage points above their expected scores in reading. Put another way, for every 100 South Shore 3rd graders, an additional 12 students met or exceeded the 3rd grade math benchmark and an additional 5 students met or exceeded the reading benchmark than would be expected had these students attended a different school.

The effects of South Shore vary as students progress through the school's grades, but South Shore students generally perform better than expected. In math, South Shore students score better than expected in all grades except 6th. In reading, South Shore students score near expectation in all grades except 3rd and 8th.

To date, South Shore has had too few 9th grade students to allow us to confidently analyze the effects of South Shore attendance beyond 8th grade. Also, it is generally more difficult to interpret an analysis for later grades because a significant number of older students transferred into South Shore in later grades and thus did not participate in many of the programs offered by South Shore. As full cohorts continue to age through the South Shore program, their performance data will illuminate the longer-term impacts of South Shore attendance.

Table 1: South Shore effects on math and reading scores, and predicted and actual shares of South Shore students meeting state standards, by grade

Math at South Shore School

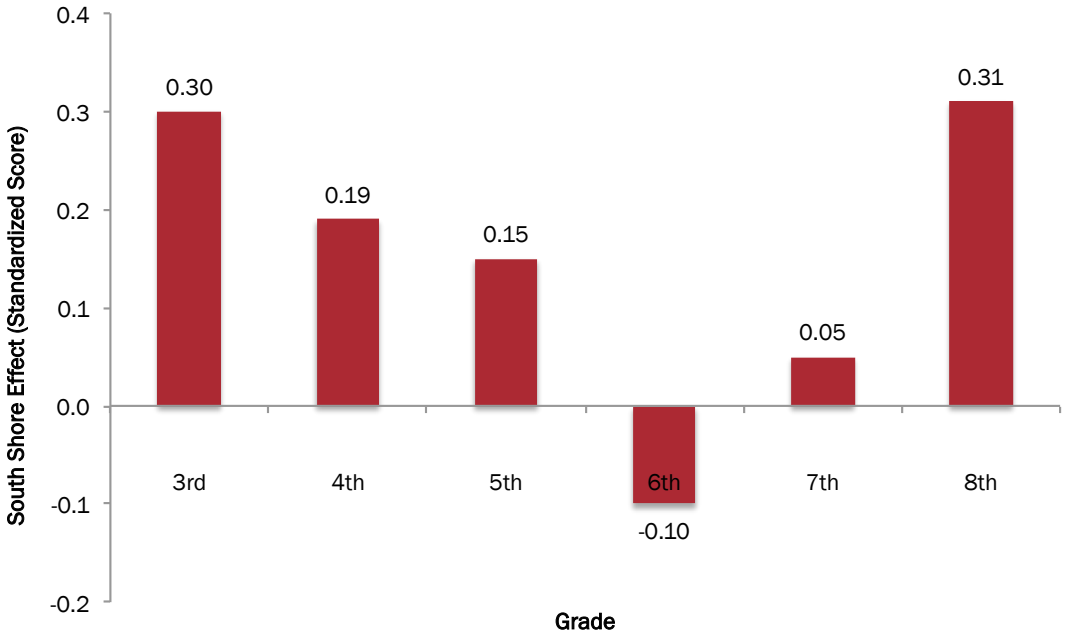
	3rd	4th	5th	6th	7th	8th
Standardized score difference	0.30	0.19	0.15	-0.10	0.05	0.31
Predicted meet rate	61%	51%	56%	60%	58%	53%
Actual meet rate	73%	62%	56%	46%	62%	58%
Difference (percentage points)	12	11	0	-14	4	5

Reading at South Shore School

	3rd	4th	5th	6th	7th	8th
Standardized score difference	0.11	-0.03	-0.03	0.02	0.04	0.37
Predicted meet rate	67%	65%	65%	70%	59%	71%
Actual meet rate	72%	66%	61%	68%	60%	79%
Difference (percentage points)	5	1	-4	-2	1	8

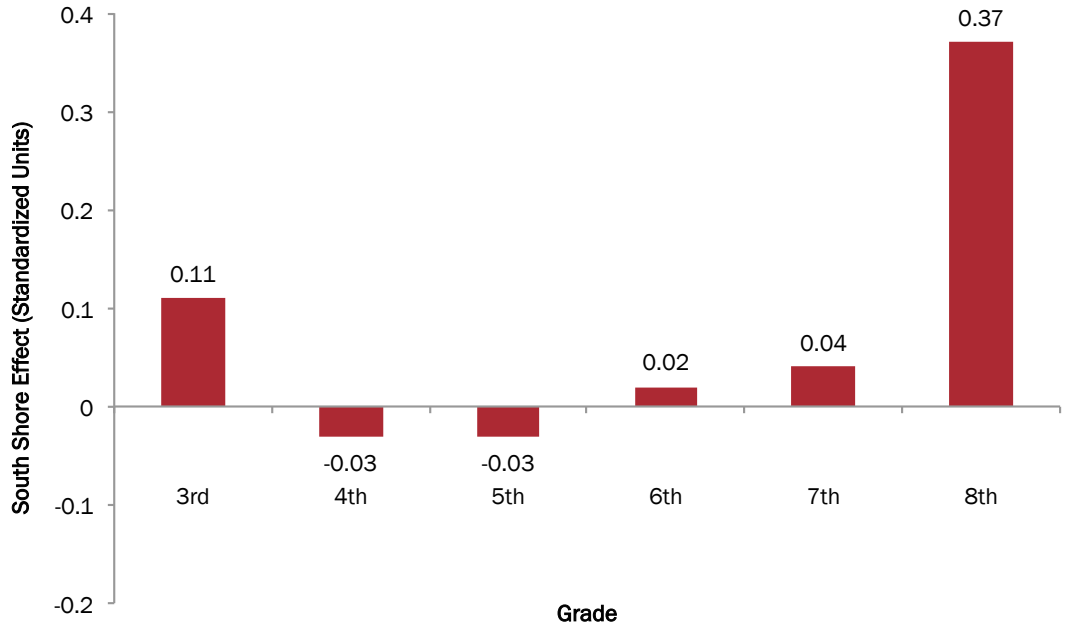
Note: Standardized score differences (South Shore effects) are the deviation from predicted performance, accounting for individual characteristics. Meet rate is the share of students meeting the state proficiency standard.
Source: ECONorthwest analysis of Seattle Public Schools data.

Figure 3: South Shore effects on math scores, by grade (deviation from expected performance, accounting for individual characteristics)



Source: ECONorthwest analysis of Seattle Public Schools data.

Figure 4: South Shore effects on reading scores, by grade (deviation from expected performance, accounting for individual characteristics)



Source: ECONorthwest analysis of Seattle Public Schools data.

The effects of starting at South Shore in Pre-K or Kindergarten

A key component of South Shore is high-quality pre-Kindergarten and Kindergarten. In this section, we examine the effects of these programs. Unfortunately, we do not currently have a reliable measure of student performance prior to third grade. Instead, we look at the effects of the pre-Kindergarten and Kindergarten programs by examining WASL and MSP scores for 3rd graders (the first grade level with test score data) by number of years attending South Shore.

Figure 5 summarizes the results of our analyses. Students who enrolled at South Shore in pre-Kindergarten or Kindergarten significantly outperformed demographically similar students attending other SPS schools in math and moderately outperformed similar students at other schools in reading. The strongest effect is for students who enrolled in the South Shore pre-Kindergarten program: math scores are 0.39 SU higher than scores for all SPS 3rd graders with similar consistent enrollment in a single school. The strongest South Shore reading effect is for students who enrolled during Kindergarten: these reading scores are 0.18 SU higher than scores for all SPS 3rd graders with similar enrollment characteristics.

These results indicate that the effects of attending South Shore are larger for students who enrolled during pre-Kindergarten or Kindergarten. While this pattern of results does not conclusively prove that South Shore's pre-Kindergarten and Kindergarten programs are responsible for these differences (the differences could reflect other unobserved differences between early and late enrollees), our results are certainly consistent with the belief that attending South Shore's pre-Kindergarten and Kindergarten programs significantly improves subsequent student performance.

The effect of South Shore on subpopulations

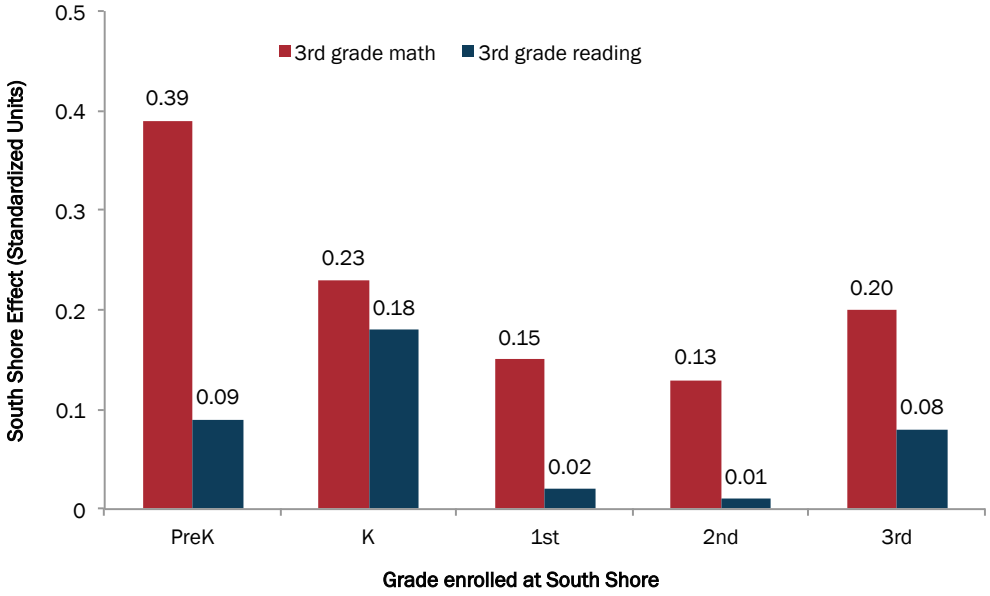
As discussed previously, South Shore attendance increases students' average scores and the likelihood of meeting the state assessment benchmarks. An important question is whether that effect is the same for all groups of students, or if it has greater impacts on some groups compared to others.

Throughout SPS, African-American students score below non-African-American students in both reading and math. The pattern generally holds at South Shore as well (although the exact magnitudes of the differences are not identical). However, relative to demographically similar African-American students attending other schools, African-American students attending South Shore have higher average scores and are more likely to meet the state benchmarks in math and reading. For instance, in 3rd grade, South Shore African-Americans score 0.26 SU higher than demographically similar African-Americans at other SPS schools in math and 0.14 SU higher in reading. These effects are roughly in line with overall differences between South Shore 3rd graders and non-South Shore 3rd graders described in the previous section.

Figure 6 shows estimates for South Shore effects on 3rd grade math scores for various subgroups. Effect sizes vary, but they are all positive and not substantially different than one another (especially given the small

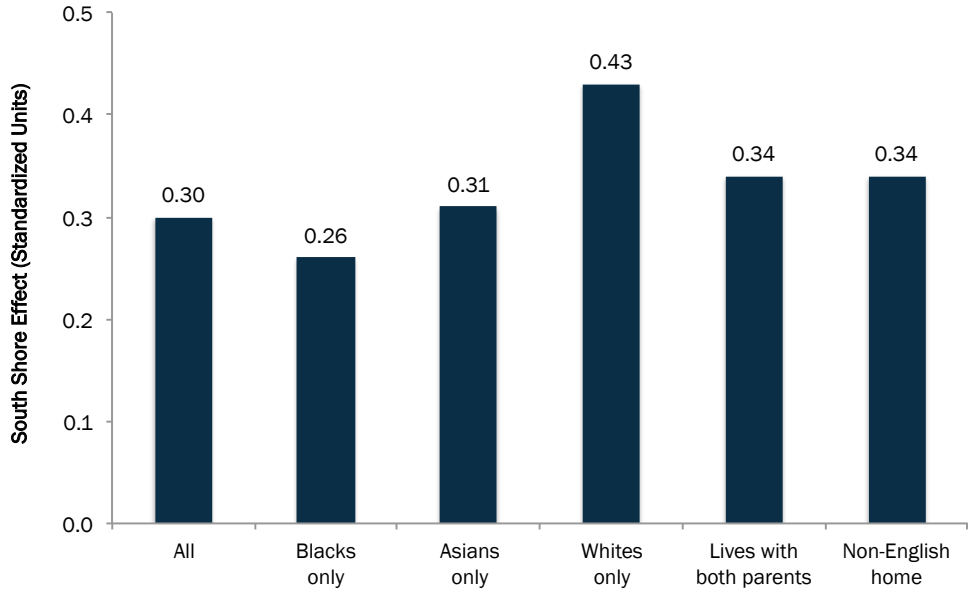
sample sizes for some groups). Among the subpopulations depicted here, white 3rd graders at South Shore have the largest estimated South Shore effect, scoring 0.43 SU higher than demographically similar white students at other SPS schools.

Figure 5: South Shore effect on math and reading scores, by students' earliest grade of enrollment at South Shore



Note: South Shore effect is the deviation from expected performance, accounting for individual student characteristics.
Source: ECONorthwest analysis of Seattle Public Schools data.

Figure 6: South Shore effects on 3rd grade math scores for various subpopulations



Source: ECONorthwest analysis of Seattle Public Schools data.

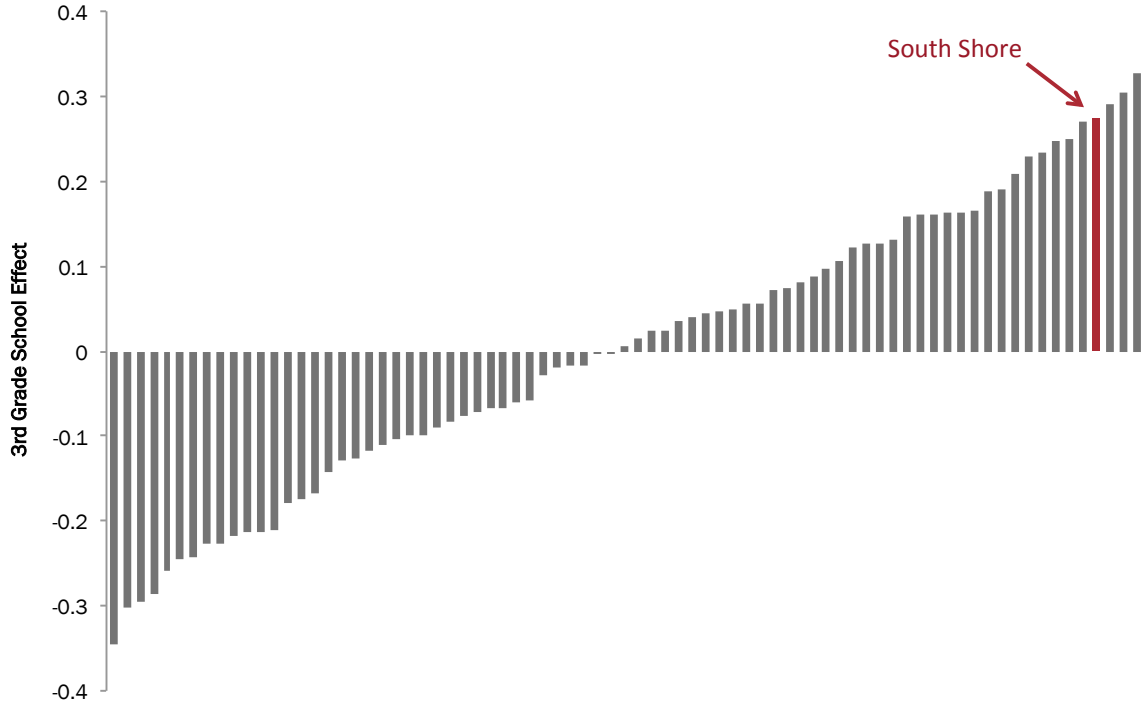
South Shore compared to other SPS elementary schools

To provide context for the South Shore effect sizes, ECONorthwest conducted similar analyses for each SPS elementary school and estimated a school effect size for each school. As described in greater detail in Appendix B, one can compute such effects in a variety of ways, but regardless of approach, South Shore ranks highly.

For 3rd grade math, the South Shore effect ranks between 1st and 6th (out of 76 schools). For 3rd grade reading, the South Shore effect ranks between 1st and 26th. As an example, Figure 7 depicts one set of rankings for 3rd grade math, where South Shore is ranked 4th. As described in Appendix B, this range in rankings reflects differences in what types of school-level characteristics are included in the calculation. Each school's estimated effects and rankings are sensitive to the outcome examined and control variables included in the analysis.

Finally, the differences in the estimated effect size among most of the other "top" schools are not statistically significant; that is, we cannot say with confidence that differences between closely ranked schools do not reflect randomness in the data. Thus, while we can reasonably assert that schools with large positive effects are improving student outcomes, we are less comfortable asserting that there is a meaningful difference between the effects of South Shore in particular and other schools with similar estimated effect sizes.

Figure 7: Distribution of SPS school effects for 3rd grade math scores (deviation from expected performance, given observable student characteristics)



Note: School effects for this particular figure are random effects estimates for a regression similar to the one described in the text, but including a school random effect for each school. See Appendix B for additional detail.
Source: ECONorthwest analysis of Seattle Public Schools data.

Conclusion

Examining student-level test data across multiple school years, ECONorthwest finds that attendance at South Shore is generally associated with improved scores on the WASL and MSP assessments for grades 3-8 and an increased likelihood of meeting the state standards in math and reading. The South Shore effects are large and statistically meaningful, particularly in math.

These effects are generally consistent across subgroups. In addition, students who attend South Shore in pre-Kindergarten and Kindergarten perform better than students who enroll at South Shore in later grades and better than their demographically similar SPS counterparts, particularly in math.

[1] South Shore's pre-Kindergarten and Kindergarten program draws on the High/Scope model, the most frequently cited demonstration of the effects of early childhood development programs. Findings from a controlled experiment, which operated in Michigan in the 1960s and tracked enrollees into adulthood, indicated that High/Scope attendees had higher high school completion rates than their non-program counterparts and outperformed the non-program group on school achievement and literacy tests at various ages. At age 40, 76 percent of program students were employed compared to 62 percent of their non-program peers. See Schweinhart, L. *Benefits, Costs, and Explanation of the HighScope Perry Preschool Program*. Paper presented at the Meeting of the Society for Research in Child Development, Tampa, Florida, April 26, 2003.

http://www.highscope.org/Content.asp?ContentId=219
http://evidencebasedprograms.org/1366-2/65-2

South Shore built its class size policy on findings from the Tennessee STAR experiment—the only wide-scale, randomized field study to examine the effects of class size on achievement. The Tennessee STAR project reduced class sizes in Kindergarten through grade 3 from about 24 students down to 15. Students in STAR's small classes performed better than students in regular classes in all locations and at every grade level. See Boyd-Zaharias, J. (1999). *Project STAR: The Tennessee Student/Teacher Achievement Ratio Study: Background and 1999 Update*. HEROS, Inc. Lebanon, TN.

[2] Kane, T., & Staiger, D.O. (2002). "Volatility in School Test Scores: Implications for Test-Based Accountability Systems" in Diane Ravitch (ed.) Brookings Papers on Education Policy, Washington, DC.

[3] This statement assumes that there would be sufficient numbers of observations in the data to prevent a handful of extreme or odd observations to be randomly assigned to one group or the other and create meaningful differences between the groups.

[4] Regression analysis is a quantitative analysis technique that describes the statistical relationship between a given variable (the dependent variable) and one or more other variables (the independent variables). For example, achievement test scores (the dependent variable in this case) might be related to certain observable characteristics such as socioeconomic status, English as a Second Language status, special education status, and in the case of this evaluation, attendance at South Shore. With a sufficient number of observations, regression analysis allows us to estimate how students included in the model who are similar in all other observable characteristics (the independent variables) will differ based on their exposure to South Shore.

[5] OSPI Assessment Overview, http://www.k12.wa.us/assessment/

[6] What Works Clearinghouse,
http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf

[7] The specific list of control variables are outlined in the data section above. Each of the discussed variables is included, untransformed, in a standard OLS regression with standard errors clustered on school (to account for the fact that students are nested within schools). In Appendix B, we present the results from alternative estimation approaches to accounting for the multi-level nature of the data, e.g., school fixed or random effects. Also, we use logistic regression to analyze outcomes with a binary outcome (e.g., whether a student met the state benchmark).

Appendix A: Assembling and Organizing the Data

Preparing the data for analysis consisted of four relatively straightforward steps. First, we converted the raw data we received into STATA (a common statistical software) format. We received comma-delimited files from SPS, and we simply read these files into STATA. We received SAS (another common statistical software) data from ERDC, and we used a program called StatTransfer to convert these files to STATA format.

Second, we removed extraneous information from each file. To reduce the size of the data files and increase processing speed and work efficiency, we identified variables in the data that were redundant or not useful for this analysis and removed them from the data (e.g., all the SPS files had variables for first and last name even though all observations has been given a "null" value).

Third, we created several new variables using the information from existing variables. Often, the data we received presented information in a text format (e.g., a test score variable was coded as text because some scores were assigned a "NULL" value). Because statistical analysis requires numeric variables, we created several variables that converted information stored as text into numbers (e.g., a gender variable that took the values "male" or "female" became a female indicator variable that took a value of 1 if gender was female and 0 if gender was male). We also converted any date variables into STATA's date format. Finally, we transformed some variables into formats to facilitate our statistical analysis. For instance, we standardized test scores by subtracting the mean outcome and dividing by the standard deviation. A standardized test score describes the number of standard deviations each student is from the mean student. This standardized measure is used commonly in the literature on school effects and it makes interpreting the magnitude of effects easier. We also created measures for growth rates (by comparing student performance across years), attendance, and suspension rates (by dividing days attended or days suspended by days enrolled).

Fourth, we reorganized the data so that each student had only one observation for each school year. Perhaps the most challenging aspect of working with student data stems from the fact that each student record does not contain precisely the same information. In particular, some students attended multiple schools per year or take multiple tests in the same subject. While we retain all of this information in our database, we only want to use a single observation for each student in our analyses. As such, we assign each student to a primary school (e.g., the school they attended the longest), and we select a single test score for each test (generally, we use the spring test attempt; however, we occasionally use alternative tests).

In sum, we gathered the following student-level data for South Shore and SPS students from 2002-03 through 2011-12:

- All test scores
- Suspensions and expulsions
- Schools attended
- School choice
- Number of days enrolled in each school
- Number of days attending each school
- Credits/GPA (middle and high school students only)
- Grade enrolled
- Ethnicity
- Sex
- English language leaner (ELL) status
- Language spoken at home
- IEP status
- Gifted status
- Living with (e.g., living with both parents)
- Free and reduced price lunch (FRL) status

Because we received these data from two sources, we created two databases to work with over the course of the project: one for SPS data and one for ERDC data. Each of these databases consists of single observations for each student in each school year, with data on student outcomes, performance, and background characteristics for each observation. The databases contain most of the same information, with two essential differences for the purposes of our evaluation: the SPS data includes school choice information, and the ERDC data includes FRL status. The school choice data from SPS allowed us to identify a set of students (and schools) that form a reasonable comparison group for South Shore students. The FRL data from ERDC allowed us to control for socioeconomic status in our analysis.

Appendix B: Estimating the South Shore Effect

Approaches to estimating the South Shore effect

The analyses described in the main body of this report required a variety of choices and assumptions with respect to how we organized our data and conducted our analyses. In this appendix, we describe our choices in greater detail and describe how the results described in the main report change when alternative choices and assumptions are employed.

Our primary purpose in conducting this analysis was to answer the question, "what is the effect of attending South Shore?" Ideally, to answer this question, we would observe outcomes for the students who attend South Shore and compare them to the outcome that these students would have obtained had they attended some other school. Of course, we cannot observe the outcomes that South Shore students would have obtained had they attended some other school. As such, we must use data on other students to construct a statistical expectation of the outcomes for each South Shore student had they not attended South Shore.

When constructing our analysis, we faced choices in N major areas:

(1) What populations to compare?
(2) What outcomes to examine?
(3) How to measure/include someone in the South Shore group?
(4) What regression specification to use to measure the South Shore effect?

Population

To calculate the South Shore effect, we compare members of the South Shore group to other students, but which other students should be included?

A highly convincing evaluation would factor out all potential differences between South Shore and non-South Shore students. To the extent that the observable variables in the data do not adequately capture all of the differences between South Shore and other schools, restricting the population may help reduce the affect of unobserved, but omitted, characteristics. As such, we use three different comparison groups: (1) all SPS students, (2) all students who ranked South Shore first in their school choice applications (thus showing similar family motivation), and (3) all students who attend schools with a relatively high transfer flow to or from South Shore.

The results from the first comparison describe the difference between the average South Shore student and the average SPS student with similar demographic characteristics.

The results from the second comparison describe the differences between the average South Shore applicant who attended South Shore and the average South Shore applicant who did not attend. If South Shore applications were denied randomly, this comparison group would be

ideal; however, applications are not denied randomly. In general, students who are admitted to South Shore are more likely to have siblings who attend and/or live closer to the school than students who apply but are not admitted.[1]

The results from the third comparison describe the difference between the average South Shore student and the average student in schools demographically more similar and geographically closer to South Shore. The schools that South Shore students are most likely to transfer to or transfer from are generally located within the same area of the city and tend to be demographically more similar to South Shore than other SPS schools.

Outcomes

As described in the main text, we examine two main outcomes: standardized test scores and met state benchmarks. When estimating equations with a binary dependent variable (i.e., met state benchmarks), we use logistic regression and report odds ratios.

Measuring South Shore

Attending South Shore does not imply a singular consistent treatment. Students who attend South Shore experience a different mix of South Shore interventions depending on when they enroll. As such, a single "average" South Shore effect is slightly misleading. While it is still informative to understand how the average South Shore student compares to a demographically similar student in some relevant comparison population, the simple average does not provide a meaningful estimate of the effectiveness of the various "extra" programs that separate South Shore from other schools. Thus we also include specifications with a separate indicator for the student's enrollment grade. In these specifications, we measure the average effect of attending South Shore starting in grade X.

Regression Specification

We use regression analysis to conduct the various analyses in this report. Specifically, we estimate equations with the following form[2]:

---

[1] This allocation mechanism raises two potential concerns for this method of evaluation. First, are applicant families who live closer to South Shore unobservably different than families living farther away in some unobservable way? For instance, does the neighborhood around South Shore contain a higher proportion of families who are more concerned about education than otherwise similar neighborhoods? Second, if they had lived farther from South Shore, would the admitted families still have applied? Because using applicants as a comparison group is supposed to account for differences in motivation to attend South Shore, we need to assume that the families who live close and

[2] All equations were estimated in STATA 12.1. When estimating equations with the binary met state benchmark outcome, we estimate the analogous logistic equations. For regressions without school random or fixed effects, we compute cluster robust standard errors. When estimating equations with fixed effects, we use STATA's areg command and recover the school fixed effects using the predict, d command. When estimating equations with random errors, we use STATA's xtmixed command with the mle option and recover the random effects using the predict, reffects command (and the standard errors for the random effects using the predict, rese command).

$$Score_{ijt} = \alpha + \beta SouthShore_{it} + X_i\delta + e_{ijt}$$

where South Shore is a variable that equals one if a student attended South Shore in year t and equals zero otherwise, and X is a vector of student demographics (gender, race, etc.). The coefficient β describes the difference in student performance between the average South Shore student and the average SPS student, controlling for various student characteristics (X). To account for the fact that students are nested within schools (and thus the error terms from this regression should not be assumed to be independent from each other), we calculate cluster-robust standard errors, clustered on school.

Generally, education researchers use a more explicit multi-level model when examining school effects. Specifically, they estimate a regression with random effects, such as:

$$Score_{ijt} = X_i\delta + Z_j\theta + \zeta_j + e_{it}$$

where X is a vector of student characteristics, Z is a vector of school characteristics, and ζ_j is a school random effect. Alternatively, researchers estimate a similar regression with school fixed effects instead of school random effects, such as:

$$Score_{ijt} = X_i\delta + \zeta_j + e_{it}$$

where the variables are the same as above, except ζ_j is a school fixed effect. In theory, the three approaches should give fairly similar answers, and they do; however, the differences in assumptions and approach to estimation do generate differences in the specific estimate for the "South Shore effect."

The biggest difference among the various approaches is the ability to include school-level controls in the random effects approach. As we show more explicitly below, the inclusion of school-level covariates (e.g., the share of students who are black or living with both parents) generally increases the estimates for the effect of attending South Shore.

Results

Table B1 presents estimates for the South Shore effect for each grade, for four different empirical specifications and three different comparison populations. While the results in the table vary some, the pattern of results is generally consistent across specifications and populations. For example, math performance starts off high in the 3rd grade and declines through 6th grade before returning once again to high levels in 8th grade.

Table B2 presents results by enrollment grade for math and reading tests in all grades. Again, the results indicate that students who enroll at South Shore in Pre-K or Kindergarten perform better than students who enroll in later grades; however, there are exceptions to this pattern. In particular, the data become fairly erratic for the later grades where the sample sizes for each enrollment grade shrink and become more subject to the influence of a handful of tests.

Table B3 presents results for various subgroups for each grade for reading and math tests. Consistent with the results described in the main report, the results for the various groups examined align, indicating that attending South Shore generates similar effects for each group. Again, there are some exceptions to this pattern, and exceptions occur more frequently in later grades with smaller sample sizes.

Table B1: South Shore effects for math and reading scores in all grades, by empirical specification and comparison population

	Full Population				Substitute Schools				South Shore Applicants			
	Standard Estimate	Fixed Effects	Random Effects, No School-Level Covariates	Random Effects With School-Level Covariates	Standard Estimate	Fixed Effects	Random Effects, No School-Level Covariates	Random Effects With School-Level Covariates	Standard Estimate	Fixed Effects	Random Effects, No School-Level Covariates	Random Effects With School-Level Covariates
Math - Grade 3	0.30	0.25	0.27	0.37	0.40	0.36	0.36	0.11	0.42	0.28	0.34	-
	(0.02)		(0.04)	(0.04)	(0.03)		(0.04)	(0.04)	(0.05)		(0.04)	
Math - Grade 4	0.19	0.12	0.16	0.21	0.30	0.27	0.29	-0.03	0.25	0.17	0.18	-
	(0.03)		(0.04)	(0.04)	(0.05)		(0.04)	(0.04)	(0.06)		(0.04)	
Math - Grade 5	0.15	0.08	0.12	0.23	0.25	0.22	0.26	-0.03	0.20	0.13	0.15	-
	(0.02)		(0.04)	(0.04)	(0.04)		(0.04)	(0.04)	(0.05)		(0.05)	
Math - Grade 6	-0.10	-0.11	0.08	0.04	-0.11	-0.12	-0.04	-	-0.14	-0.08	-0.08	-
	(0.05)		(0.05)	(0.05)	(0.12)		(0.04)		(0.11)		(0.05)	
Math - Grade 7	0.05	0.04	0.07	0.13	0.06	0.05	0.09	-	-0.01	0.01	0.12	-
	(0.06)		(0.06)	(0.06)	(0.11)		(0.06)		(0.10)		(0.07)	
Math - Grade 8	0.31	0.27	0.35	0.08	0.35	0.25	0.39	-	0.36	0.11	0.18	-
	(0.07)		(0.08)	(0.08)	(0.13)		(0.09)		(0.18)		(0.09)	
Reading - Grade 3	0.11	0.05	0.08	0.16	0.17	0.15	0.16	0.06	0.16	0.12	0.15	-
	(0.02)		(0.04)	(0.04)	(0.03)		(0.04)	(0.03)	(0.05)		(0.04)	
Reading - Grade 4	-0.03	-0.08	-0.04	-0.02	0.03	0.02	0.04	-0.08	0.00	0.01	0.00	-
	(0.02)		(0.04)	(0.04)	(0.04)		(0.04)	(0.04)	(0.05)		(0.05)	
Reading - Grade 5	-0.03	-0.08	-0.05	0.03	0.03	0.03	0.05	-0.07	-0.03	-0.01	-0.01	-
	(0.02)		(0.05)	(0.04)	(0.04)		(0.05)	(0.04)	(0.05)		(0.05)	
Reading - Grade 6	0.02	0.00	0.17	-0.06	0.03	0.01	0.03	-	0.01	0.01	0.01	-
	(0.03)		(0.05)	(0.05)	(0.05)		(0.05)		(0.09)		(0.06)	
Reading - Grade 7	0.04	0.04	0.04	0.01	0.05	0.04	0.07	-	0.03	0.02	0.02	-
	(0.04)		(0.06)	(0.03)	(0.08)		(0.06)		(0.11)		(0.06)	
Reading - Grade 8	0.37	0.37	0.26	0.09	0.41	0.40	0.28	-	0.30	0.22	-	-
	(0.04)		(0.08)	(0.05)	(0.07)		(0.08)		(0.18)			

Note: Standard errors are in parentheses.
Source: ECONorthwest analysis of Seattle Public Schools data.

Table B2: South Shore effects for math and reading scores in all grades, by South Shore enrollment grade

	Enrollment Grade									
	Pre-K	K	1st	2nd	3rd	4th	5th	6th	7th	8th
Math - Grade 3	0.39	0.23	0.15	0.13	0.20					
	(0.03)	(0.05)	(0.08)	(0.16)	(0.06)					
Math - Grade 4	0.27	0.39	-0.08	0.11	-0.11	0.08				
	(0.03)	(0.05)	(0.06)	(0.07)	(0.07)	(0.07)				
Math - Grade 5	0.23	0.22	-0.05	0.02	-0.05	0.07	0.17			
	(0.03)	(0.05)	(0.06)	(0.03)	(0.11)	(0.09)	(0.08)			
Math - Grade 6	0.03	0.02	-0.30	-0.20	-0.38	-0.04	-0.12	0.09		
	(0.05)	(0.05)	(0.16)	(0.06)	(0.09)	(0.08)	(0.07)	(0.10)		
Math - Grade 7	0.16	-0.16	0.34	0.11	-0.43	0.48	-0.08	0.06	0.18	
	(0.07)	(0.07)	(0.17)	(0.18)	(0.12)	(0.17)	(0.12)	(0.10)	(0.15)	
Math - Grade 8	0.37	0.10	0.58	0.17	0.13	-0.12	0.29	0.31	-0.44	0.48
	(0.09)	(0.09)	(0.06)	(0.08)	(0.28)	(0.20)	(0.06)	(0.08)	(0.19)	0.49
Reading - Grade 3	0.09	0.18	0.02	0.01	0.08					
	(0.03)	(0.04)	(0.06)	(0.15)	(0.07)					
Reading - Grade 4	-0.03	0.12	-0.18	0.00	-0.27	0.02				
	(0.02)	(0.04)	(0.03)	(0.06)	(0.09)	(0.09)				
Reading - Grade 5	0.00	0.09	-0.12	-0.02	-0.23	0.01	-0.02			
	(0.03)	(0.07)	(0.05)	(0.04)	(0.09)	(0.07)	(0.12)			
Reading - Grade 6	-0.02	0.25	0.15	-0.03	-0.14	0.26	-0.21	0.08		
	(0.03)	(0.05)	(0.09)	(0.09)	(0.13)	(0.15)	(0.06)	(0.07)		
Reading - Grade 7	-0.01	0.32	0.52	0.13	-0.24	0.78	-0.03	-0.05	0.24	
	(0.04)	(0.04)	(0.12)	(0.10)	(0.09)	(0.12)	(0.23)	(0.02)	(0.16)	
Reading - Grade 8	0.45	0.51	0.48	-0.24	-0.30	0.51	0.55	0.38	-0.31	0.41
	(0.04)	(0.04)	(0.03)	(0.05)	(0.10)	(0.10)	(0.41)	(0.05)	(0.10)	(0.18)

Note: Standard errors are in parentheses.
Source: ECONorthwest analysis of Seattle Public Schools data.

Table B3: South Shore effects for math and reading scores in all grades, by student characteristics

	Black only	Asian only	White only	Living with both parents	Non-English Home
Math - Grade 3	0.26	0.31	0.43	0.34	0.34
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Math - Grade 4	0.20	0.26	0.16	0.20	0.27
	(0.04)	(0.04)	(0.02)	(0.03)	(0.04)
Math - Grade 5	0.15	0.21	0.12	0.10	0.13
	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
Math - Grade 6	-0.11	-0.14	0.04	-0.40	-0.11
	(0.07)	(0.08)	(0.04)	(0.06)	(0.06)
Math - Grade 7	0.22	0.04	0.28	-0.18	0.07
	(0.07)	(0.06)	(0.07)	(0.05)	(0.06)
Math - Grade 8	0.17	0.14	0.68	-0.01	-0.07
	(0.10)	(0.07)	(0.09)	(0.07)	(0.09)
Reading - Grade 3	0.14	0.14	0.13	0.07	0.14
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Reading - Grade 4	0.00	0.07	0.03	-0.06	0.00
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
Reading - Grade 5	0.00	0.06	-0.08	-0.04	-0.02
	(0.03)	(0.04)	(0.02)	(0.02)	(0.04)
Reading - Grade 6	0.00	-0.01	0.20	0.08	0.08
	(0.05)	(0.05)	(0.04)	(0.04)	(0.06)
Reading - Grade 7	0.27	-0.11	0.33	-0.26	0.08
	(0.06)	(0.07)	(0.04)	(0.04)	(0.06)
Reading - Grade 8	0.46	0.27	0.65	0.05	0.27
	(0.09)	(0.04)	(0.04)	(0.03)	(0.04)

Note: Standard errors are in parentheses.
Source: ECONorthwest analysis of Seattle Public Schools data.