VIF Splash Dual Language Immersion Program:

An Evaluation of Student Outcomes

2010-11 through 2012-13

December, 2014

Prepared for:

VIF International Education

Prepared by:

Kristina M. Patterson

Education Policy Initiative at Carolina,

University of North Carolina at Chapel Hill

kpatt@unc.edu

Executive Summary

To extend its commitment to authentic language learning, VIF International Education established the Splash dual language immersion program in 2006. The K-5 Splash immersion program offers instruction in the core academic content areas in both English and the target immersion language. The structure of Splash program varies based on the needs of the school community. Most of the programs are one-way immersion programs which target English speaking students, and about a quarter of the programs are two-way programs which are aimed at native English speakers as well as native speakers of the target language. Students in the Splash program are expected to develop proficiency in both their first language and the target language, develop positive cross-cultural behaviors, and to be at or above grade level in terms of academic achievement. Currently, VIF partners with 20 school districts to offer immersion programs in both Spanish and Mandarin in 42 elementary schools and extension programs in 3 middle schools.

In order to assess the impact of the Splash program on participants, this study analyzes data provided by the Education Policy Initiative at Carolina with permission from the University of North Carolina General Administration, to address the following research questions:

1. How do Splash dual language immersion students compare to non-Splash students in the same schools in terms of:
    a. Demographic characteristics,
    b. Achievement, as measured by test score outcomes, and
    c. Attendance?
2. How does the performance of Splash students vary by subgroup, i.e. students identified as economically disadvantaged, exceptional children, limited English proficient, or racial and ethnic minorities?

The following report outlines the findings from the study, comparing Splash students and non-Splash students who attend the same elementary schools. We found some key demographic differences between the two groups of students and these differences vary between schools. Overall, males are underrepresented in Splash classrooms, as are students who are eligible for free and reduced price lunch and students identified as exceptional children, while students identified as academically and intellectually gifted are overrepresented in Splash classrooms as compared to non-Splash classrooms. These differences point to a need to investigate the recruitment strategies and practices of many Splash schools.

Our study indicates that, on average, students in Splash classrooms have higher scores on math, reading, and science End of Grade (EOG) tests than their peers in non-Splash classrooms and their peers across North Carolina. Students in Splash classrooms also demonstrate higher rates of proficiency on EOG tests in 3rd grade math and reading, 4th grade reading, and 5th grade math and science. We find similar trends in student subgroups. Splash students identified as limited English proficient, and those who are eligible for free and reduced price lunch also, on average, score higher on EOG tests than their non-Splash peers. Splash students identified as racial and ethnic minorities score as well or higher than their non-Splash peers on EOG tests.

As participation in the Splash program is voluntary, to control for selection into the program we constructed a matched sample based on a range of factors which are related to participation in the

Splash program as well as test score outcomes. Analysis of this matched sample indicates that participation in the Splash program has a positive impact on students' math and reading EOG test scores. There is evidence that the effect on math EOG scores may be even stronger for students in Splash classrooms who are eligible for free and reduced price lunch.

Introduction

To extend its commitment to authentic language learning, VIF International Education established the Splash dual language immersion program in 2006. The K-5 Splash immersion program offers instruction in the core academic content areas in both English and the target immersion language. The structure of Splash program varies based on the needs of the school community. Most of the programs are one-way immersion programs which target native English speaking students, and generally start with 90% of instruction in the target language. About a quarter of the programs are two-way programs which are aimed at native English speakers as well as native speakers of the target language. The two-way programs use a two teacher model--one native English speaking and the other a native speaker of the target language-- and start at 50% English and 50% target language instruction. Students in the Splash program are expected to develop proficiency in both their first language and the target language, develop cross-cultural behaviors, and to be at or above grade level in terms of academic achievement. Currently, Splash partners with 19 school districts to offer immersion programs in both Spanish and Mandarin in 42 elementary schools and extension programs in 3 middle schools.

Purpose of the Report

To better understand the impact of the VIF Splash Dual Language Immersion program on participating students in North Carolina public schools, this evaluation used quantitative analysis to answer the following research questions:

1. How do Splash schools compare to non-Splash schools in terms of school size, discipline, performance on EOG tests, and expenditures?
2. How do VIF Splash dual language immersion students compare to non-Splash students in terms of:
    a. Demographic characteristics,
    b. Achievement, as measured by test score outcomes, and
    c. Attendance?
3. How does the performance of VIF Splash Language Immersion students vary by subgroup (i.e., racial and ethnic minorities, students identified as exceptional children (EC), students identified as academically and intellectually gifted (AIG), students identified as limited English proficiency (LEP), students eligible for free and reduced price lunch(FRL))?

The evaluation was conducted at the request of VIF and the research questions were developed in conjunction with VIF staff, however, the researcher independently designed the methodology, conducted the analyses, and presented the findings, conclusions, and recommendations.

This report outlines the findings from the analyses, discusses the implications of the findings, and offers recommendations for next steps.

Methods

This study analyzes student level data from seven Splash elementary schools from six school districts, during the 2010-11, 2011-12, and 2012-13 school years. All of the schools are public elementary schools in North Carolina. Our evaluation focuses on elementary school Spanish programs

during the 2010-11, 2011-12, and 2012-13 school years. Five of the schools offer one-way programs and two schools offer two-way programs. The first cohort of Splash participants entered the 3rd grade in 2010-11, therefore, this is the first year for which student End of Grade test score data was available for Splash participants. All data was provided by the Education Policy Initiative at Carolina with permission from the University of North Carolina General Administration.

Descriptive analysis techniques were used to compare schools with the Splash program and schools that do not offer the program, as well as students who participate in the Splash program and students who do not participate. Appropriate statistical tests were conducted to determine if there were differences between the schools and the groups of students on various characteristics. In order to ensure the reliability of estimates, and to protect the identity of students, we do not report findings for any group with fewer than ten students.

Since participation in Splash is voluntary, to control for selection based on observed student characteristics, propensity score matching was used, specifically one-to-one nearest neighbor matching without replacement within calipers. Each 3rd grade Splash student was matched to one non-Splash 3rd grade student within the same school, with the most similar propensity score. If a student with a similar propensity score (within ¼ of a standard deviation) could not be found, the student was eliminated from the sample (Rosenbaum and Rubin, 1985). A number of within-study comparisons demonstrate that propensity score matching techniques offer significant bias reduction, producing estimates that are similar to random assignment studies if the variables that predict treatment are correctly and fully specified (Diaz and Handa, 2005; Cook, Shadish, and Wong, 2008; Henry and Yi, 2009).

After matching, to evaluate the impact of the Splash program on student test scores, multiple regression analysis was used, including covariates that are potentially correlated with selection into Splash and test score outcomes, as well as the propensity score in the regression equation to further reduce bias (Glazerman, Levy, and Myer, 2003). School fixed effects were used to limit the comparison to students in the same schools.

Findings

Splash Schools vs. Non-Splash Schools

As shown in Table 1, a comparison of the elementary schools that offer Splash and other elementary schools across North Carolina during the 2012-13 school year indicates that the schools are fairly similar on most school characteristics, including End of Grade (EOG) test performance, discipline and safety, student demographic makeup, and expenditures. Schools with dual language programs other than Splash were excluded from this comparison. The only significant differences between Splash schools and non-Splash elementary schools are that Splash schools tend to have larger student populations and a higher proportion of novice teachers, or teachers with three or fewer years of experience in the classroom, than schools that do not offer dual language programs.

Table 1: Splash Schools Compared to Non-DLI Schools, 2012-13

	Splash Schools N=7	Non-DLI Schools N=1348
Average Daily Membership	651*	497*
Performance Indicators		
% Met or Exceeded Expected EVAAS Growth	85.71	74.39
Overall Performance Composite	40.44	42.64
Faculty		
% Fully Licensed Teachers	98.97	98.09
% Novice Teachers (<= 3years experience)	36.09*	19.54*
% National Board Certified Teachers	13.46	17.03
% with Advanced Degrees	22.78	29.67
Discipline and Safety		
Short Term Suspension rate	5.71	8.84
Violent Acts/1000	2.19	2.89
Student Demographic Makeup		
% Qualify for Free/Reduced Price Lunch	71.49	66.11
% Asian	0.92	2.27
% Black	28.12	26.00
%Hispanic	22.62	15.56
% Multiracial	4.56	3.81
Financial		
Total per pupil expenditures	$8485	$9602
Average Teacher Salary Supplement	$2829.57	$3223.30

** indicates a statistically significant difference between Splash and other schools, p<=0.05*

Splash Students vs. Non-Splash Students

Demographics

As shown in Table 2, while Splash students and non-Splash students in the same schools are similar on many characteristics, they differ on some key demographic characteristics. Race and ethnicity, as well as identification as Limited English Proficient are distributed similarly between the two groups. As compared to students in the same schools, males are underrepresented in Splash classrooms, as are students who are eligible for free or reduced price lunch, students identified as exceptional children, and students who are overage for grade, defined as a having a birth date a full year before the date their cohort could legally start Kindergarten. Conversely, students identified as

academically or intellectually gifted are overrepresented in Splash classrooms as compared to non-Splash classrooms in the same schools.

Table 2: Splash Students Compared to Non-Splash Students in the Same Schools, 2012-13

	Splash Students N=282	Non-Splash Students in the Same Schools N=1692
Gender		
Male	39.36%*	52.07%*
Female	60.64%*	47.93%*
Race/Ethnicity		
Asian	0.71%	1.06%
Black	25.18%	27.42%
Hispanic	27.30%	23.64%
American Indian	0.35%	0.53%
Multiracial	4.26%	3.25%
Pacific Islander	0.00%	0.06%
White	42.20%	44.03%
Limited English Proficient	8.87%	12.00%
Formerly LEP	5.67%	6.26%
Free/Reduced Price Lunch	47.52%*	66.55%*
Exceptional Children	2.48%*	12.45%*
AIG	12.06%*	7.69%*
Underage for Grade	0.71%	0.67%
Overage for Grade	3.19%*	19.46%*

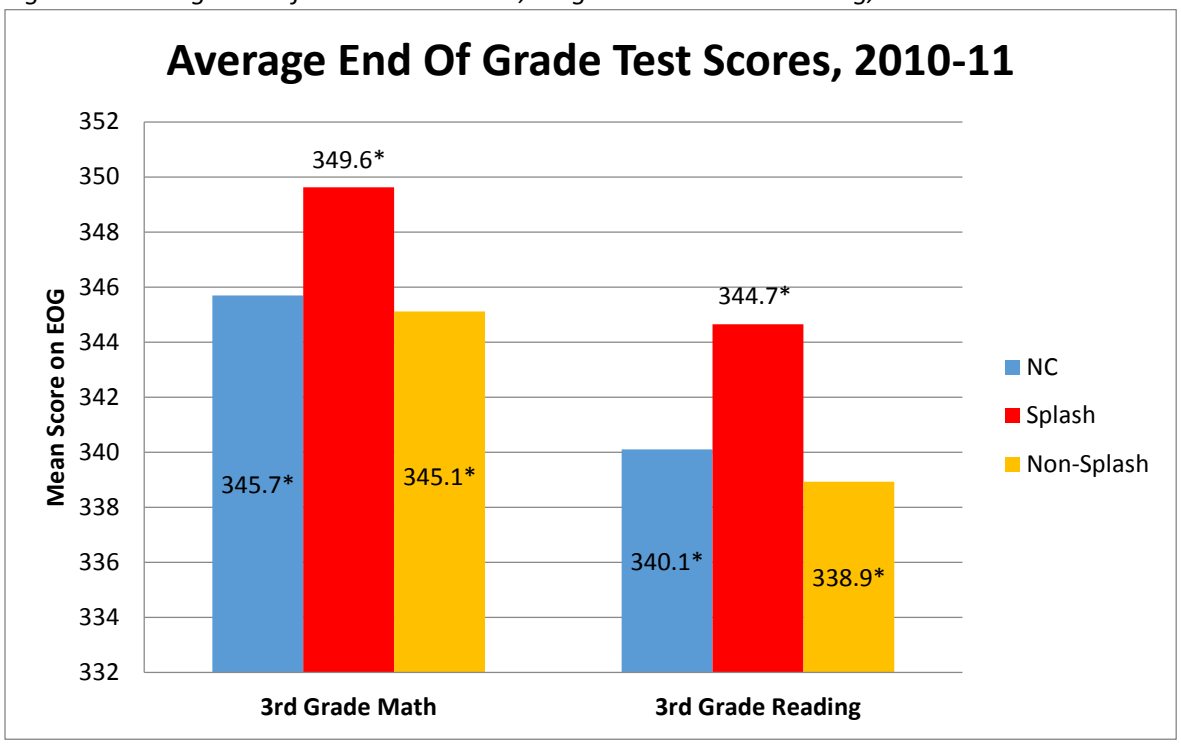
** indicates a statistically significant difference between Splash and other students, p<=0.05*

Student Achievement, as measured by End of Grade test scores

As demonstrated in Figures 1-4, on average Splash students consistently demonstrate results on NC End of Grade standardized tests that outpace their peers within the same schools and statewide. Throughout this report, we discuss those results that are statistically significant at the 95% level (p<=0.05). In 2010-11, Splash students' End of Grade test scores were higher than their peers both in the same schools and statewide in 3rd grade Math and Reading. In 2011-12, Splash students' End of Grade test scores were higher than their peers in the same schools and statewide in 3rd grade Math and Reading and 4th grade Math and Reading. In 2012-13, Splash students' End of Grade test scores were higher than their peers in the same school in all elementary school tested grades and subjects. Splash students' End of Grade test scores were significantly higher than their peers statewide in 3rd grade Math and Reading, 4th grade Math and Reading, and 5th grade Math and Science.

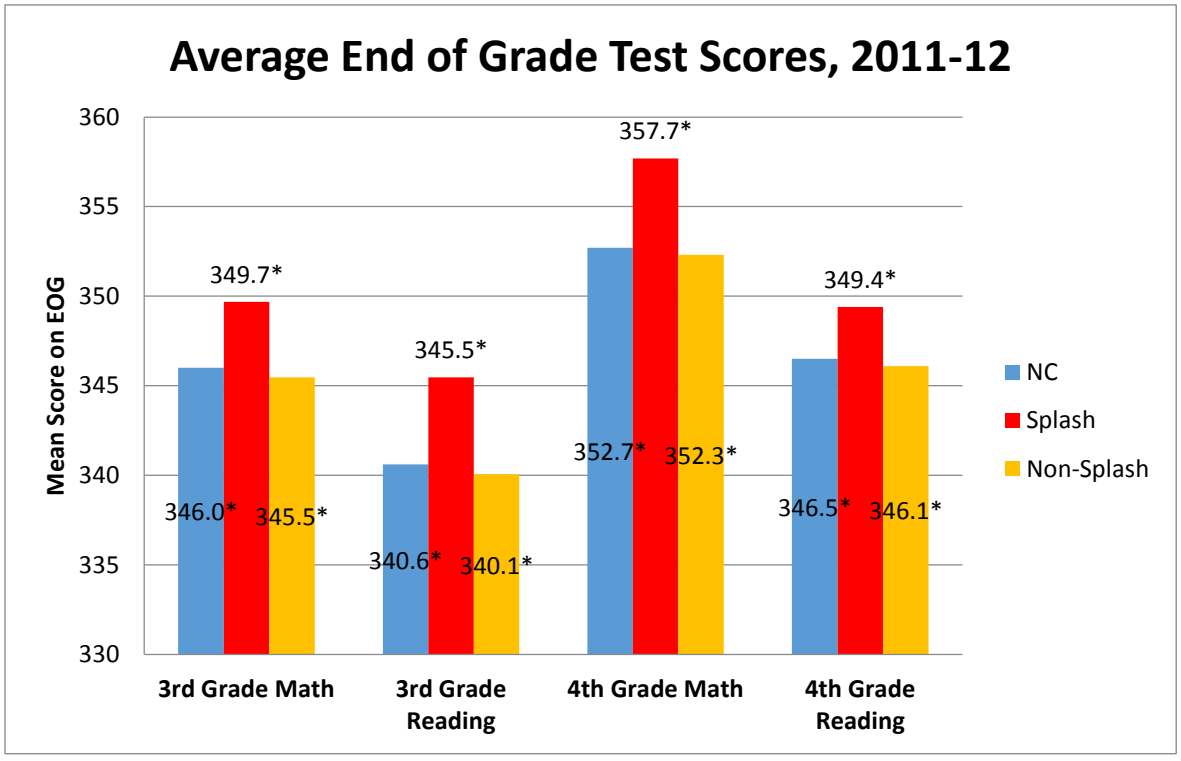
Note: North Carolina averages are from *The North Carolina State Testing Results* ("The Green Book") published by the North Carolina Department of Public Instruction. Statistical analyses have been conducted removing Splash students from the sample in order to compare scores of Splash students to students statewide.

Figure 1: Average End of Grade Test Scores, 3rd grade Math and Reading, 2010-11



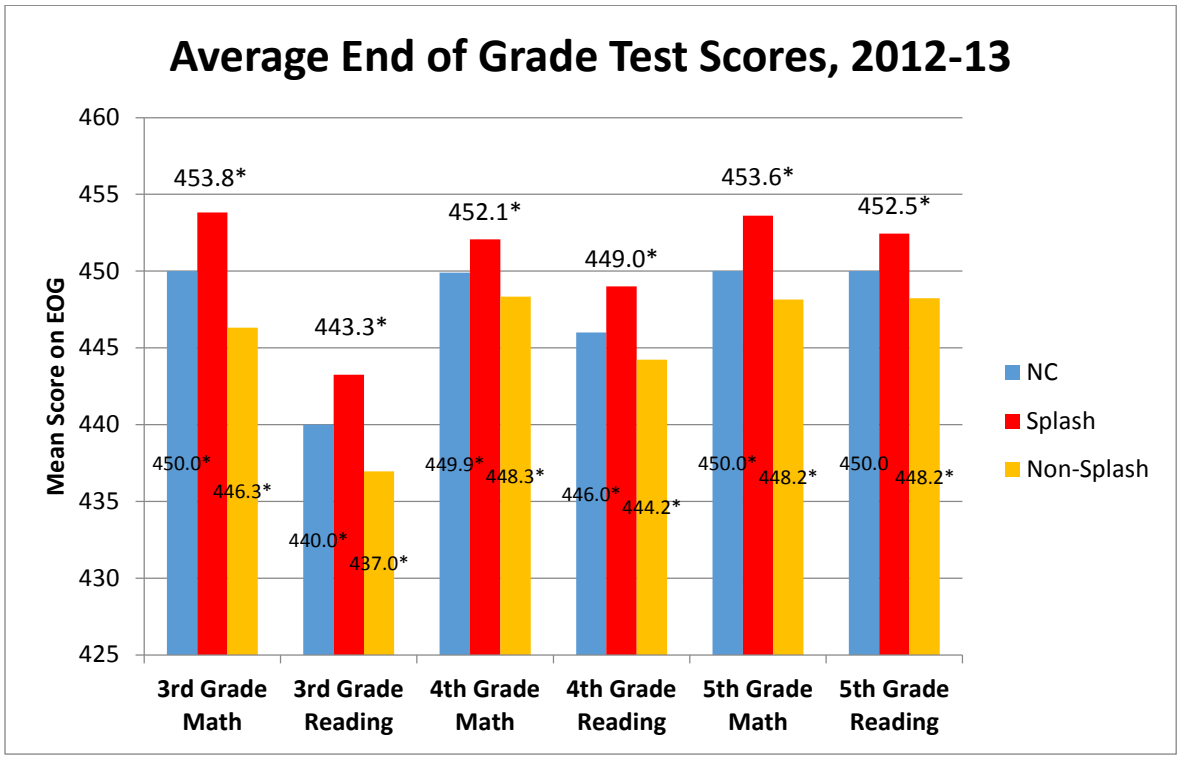
*indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05

Figure 2: Average End of Grade Test Scores, 3rd and 4th grade Math and Reading, 2011-12



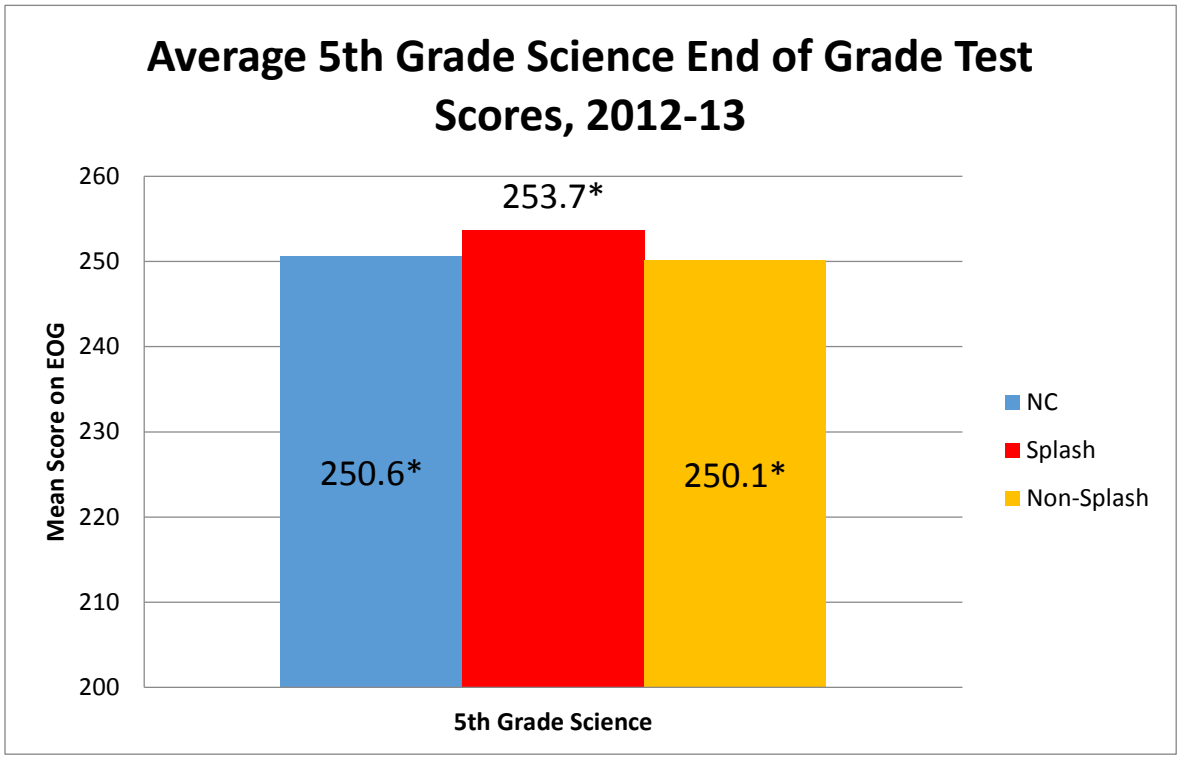
**indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05*

Figure 3: Average EOG Test Scores, 3rd, 4th, and 5th grade Math and Reading, 2012-13



*indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05

Figure 4: Average EOG Test Scores: 5th Grade Science, 2012-13

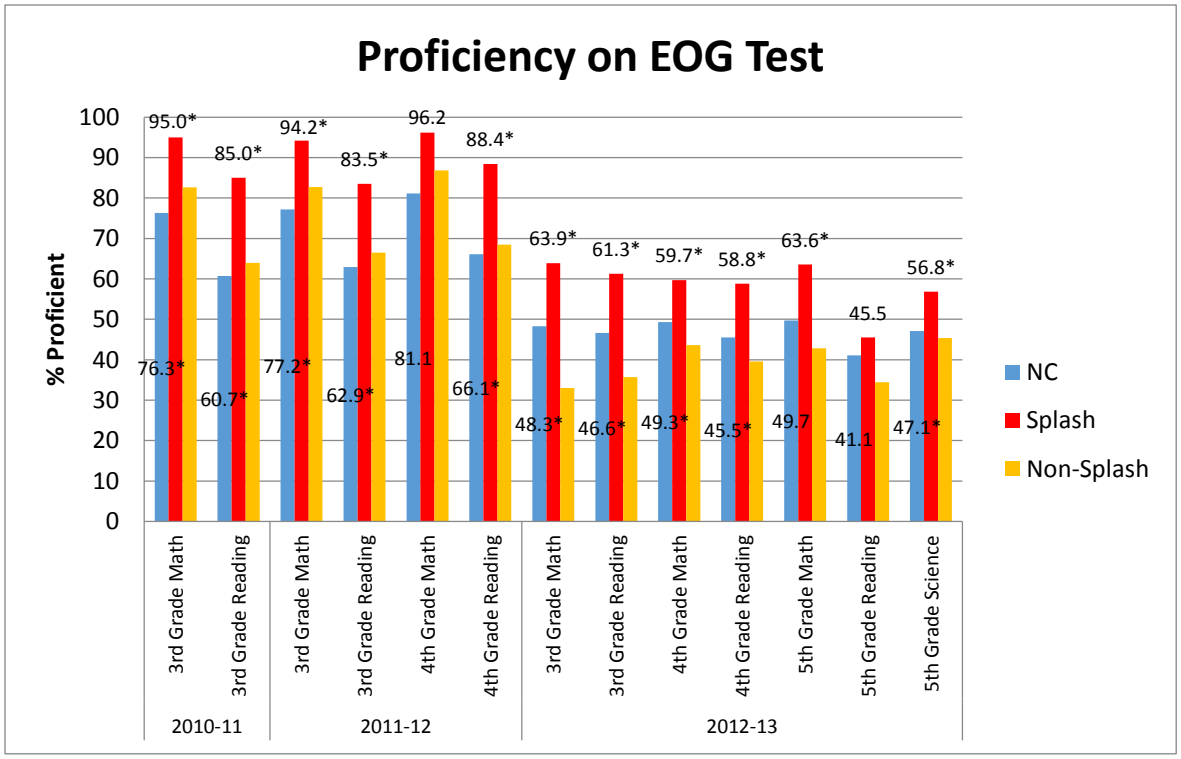


*indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05

As demonstrated in Figure 5, in addition to higher average test scores, Splash students consistently demonstrate higher rates of proficiency on End of Grade tests than their peers in the same schools and statewide. A significantly higher proportion of Splash students achieved proficiency in 3rd grade Math and Reading (2010-11, 2011-12, 2012-13), 4th grade Reading (2011-12 and 2012-13), 4th grade Math (2012-13), and 5th grade Math and Science (2012-13), than their peers in the same schools. By raising the overall proficiency rate of the school, Splash students support their schools in reaching both federal and state Annual Measurable Objectives (AMO) targets. In addition, a significantly higher proportion of Splash students demonstrated proficiency than their peers statewide in 3rd grade Math and Reading (2010-11, 2011-12, 2012-13), 4th grade Reading (2011-12 and 2012-13), 4th grade Math (2012-13), and 5th grade Science (2012-13).

It is important to note that proficiency rates decrease among all students in 2012-13 due to changes in the End of Grade tests in this year.

Figure 5: Proficiency on EOG Tests, 2010-11 through 2012-13



*indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05

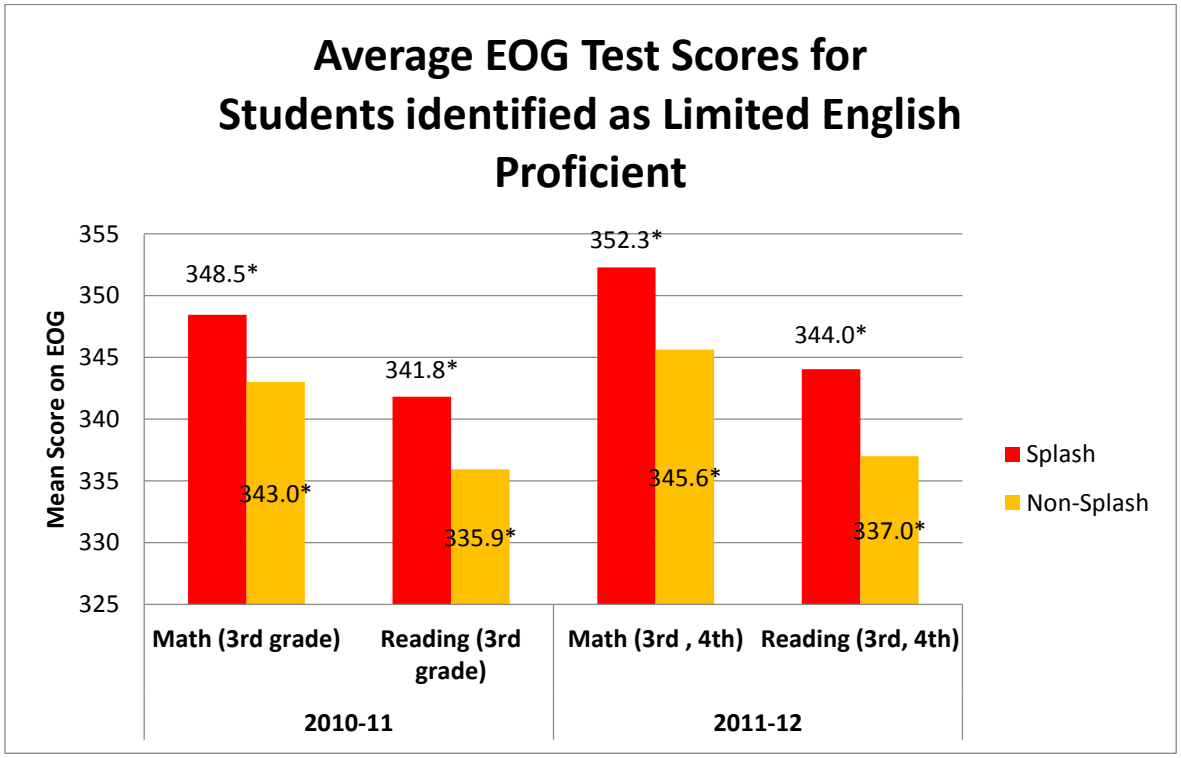
Student achievement, by subgroups

In addition to overall EOG test score performance by Splash students, we also examined EOG test scores and proficiency rates by the following subgroups: students identified as Limited English Proficient, students eligible for free and reduced price lunch, students identified as academically and intellectually gifted, and students identified as Black, Hispanic, and White. There were too few Splash students identified as exceptional children to report results on this subgroup.

Students identified as Limited English Proficient

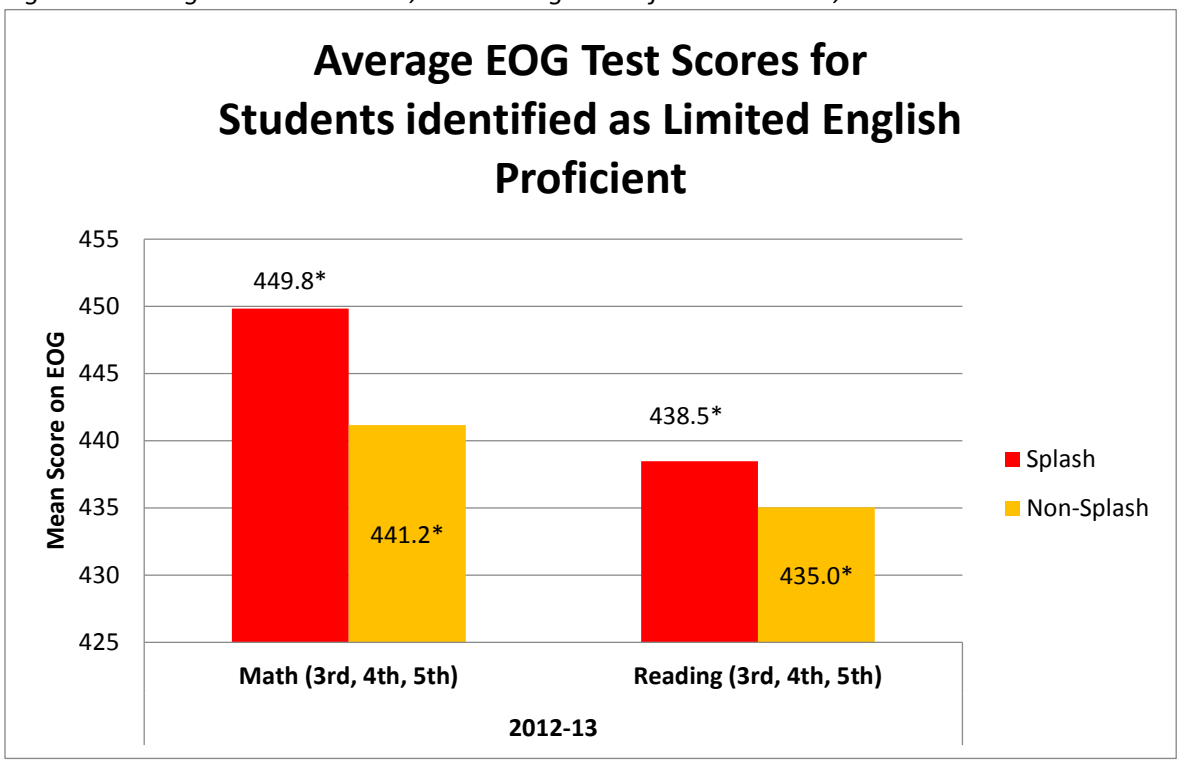
As depicted in Figures 6 and 7, Splash students identified as Limited English Proficient outperformed their peers in the same schools on 3rd grade Math and Reading EOG tests in 2010-11, 3rd and 4th grade Math and Reading EOG tests in 2011-12, and 3rd, 4th, and 5th grade Math and Reading EOG tests in 2012-13. There are too few 5th Grade Splash students identified as Limited English Proficient in 2012-13 to report performance on the 5th grade Science EOG test. In this report, average EOG test scores are not reported by subgroup to remain consistent with the North Carolina Department of Public Instruction's practices.

Figure 6: Average EOG Test Scores, Limited English Proficient Students, 2010-11, 2011-12



*indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05

Figure 7: Average EOG Test Scores, Limited English Proficient Students, 2012-13

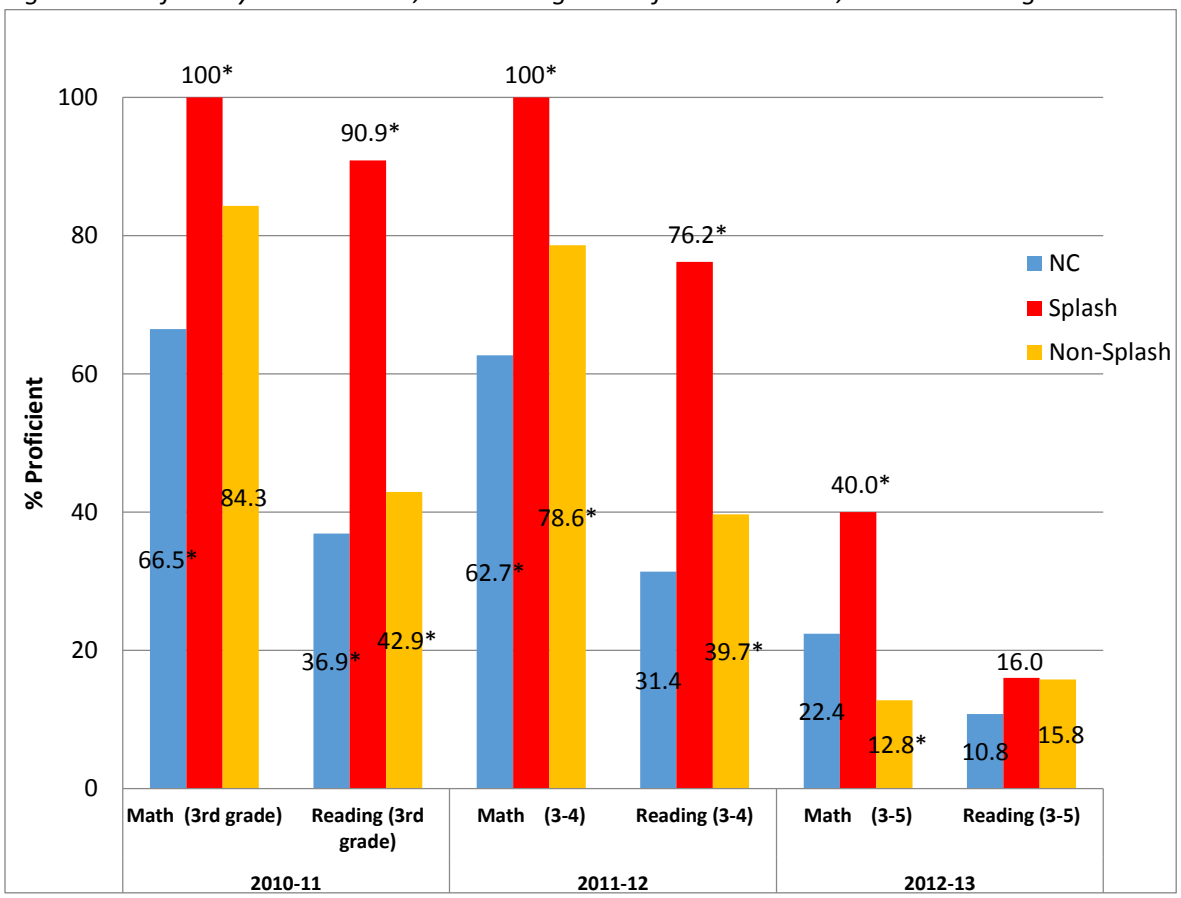


*indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05

As illustrated in Figure 8, a significantly higher proportion of Splash students identified as Limited English Proficient (LEP) demonstrated proficiency on the EOG tests in Reading in 2010-11, Reading and Math in 2011-12, and Math in 2012-13, than their LEP identified peers in the same schools. As Annual Measurable Objectives (AMO) targets include proficiency rates for students identified as Limited English Proficient, Splash students again help their schools reach both federal and state AMO targets. In addition, a significantly higher proportion of Splash students identified as Limited English Proficient demonstrated proficiency than their LEP identified peers statewide in Math in 2010-11 and 2011-12, and Reading in 2010-11. There were too few Splash students identified as Limited English Proficient with 5th grade Science EOG scores to report results for this test. Notably, 100% of LEP identified Splash students were proficient on Math EOG tests in 2010-11 and 2011-12.

As noted above, proficiency rates decrease among all students in 2012-13 due to changes in the End of Grade tests in this year.

Figure 8: Proficiency on EOG Tests, Limited English Proficient Students, 2010-11 through 2012-13

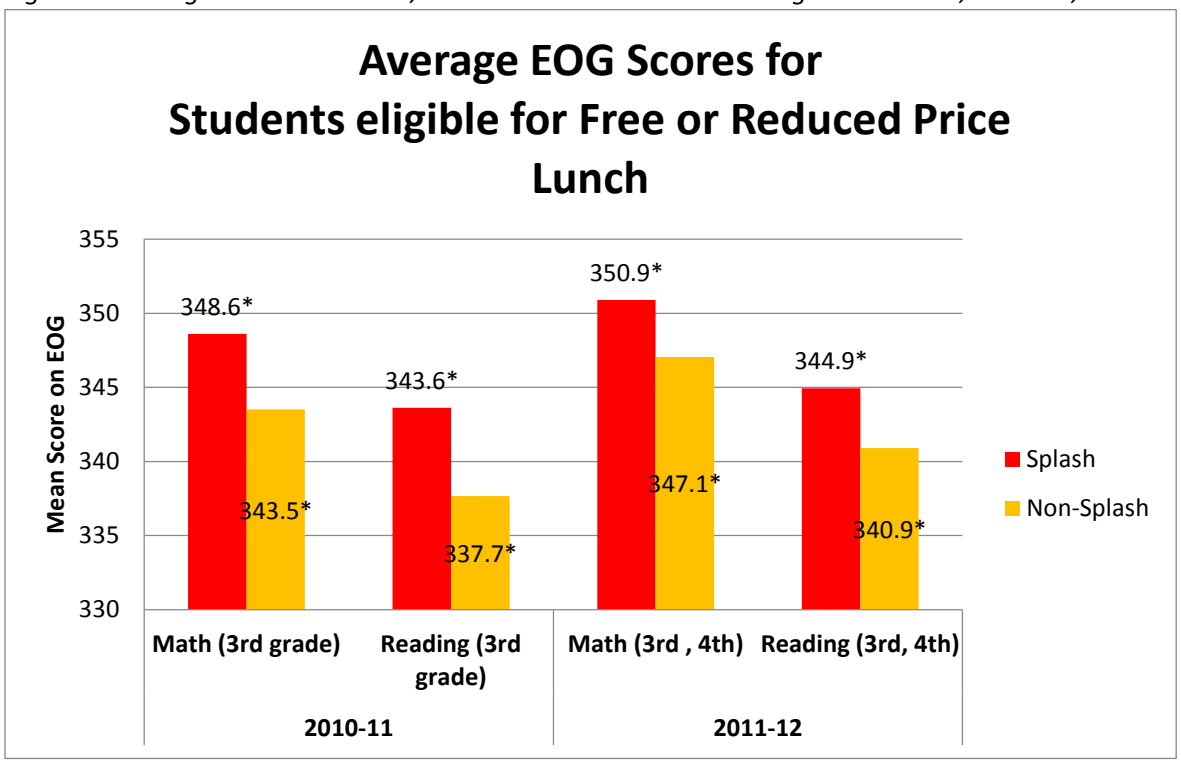


*indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05

Students Eligible for Free/Reduced Price Lunch

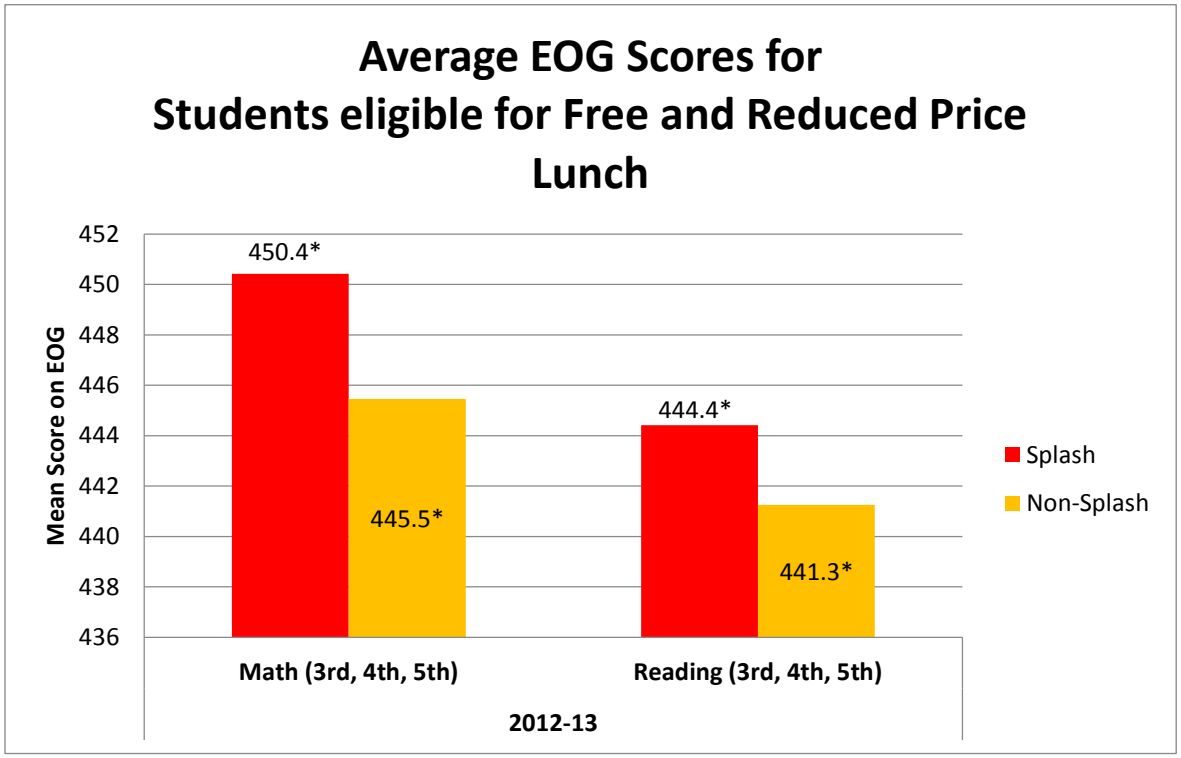
As depicted in Figures 9 and 10, Splash students eligible for free or reduced price lunch (FRL) outperformed their economically disadvantaged peers in the same schools on 3rd grade Math and Reading EOG tests in 2010-11, 3rd and 4th grade Math and Reading EOG tests in 2011-12, and 3rd, 4th, and 5th grade Math and Reading EOG tests in 2012-13. There were too few 5th Grade Splash students eligible for free or reduced price lunch to report performance on the 5th grade Science EOG test. In this report, average EOG test scores are not reported by subgroup to remain consistent with the North Carolina Department of Public Instruction's practices.

Figure 9: Average EOG Test Scores, Free or Reduced Price Lunch Eligible Students, 2010-11, 2011-12



*indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05

Figure 10: Average EOG Test Scores, Free or Reduced Price Lunch Eligible Students, 2012-13

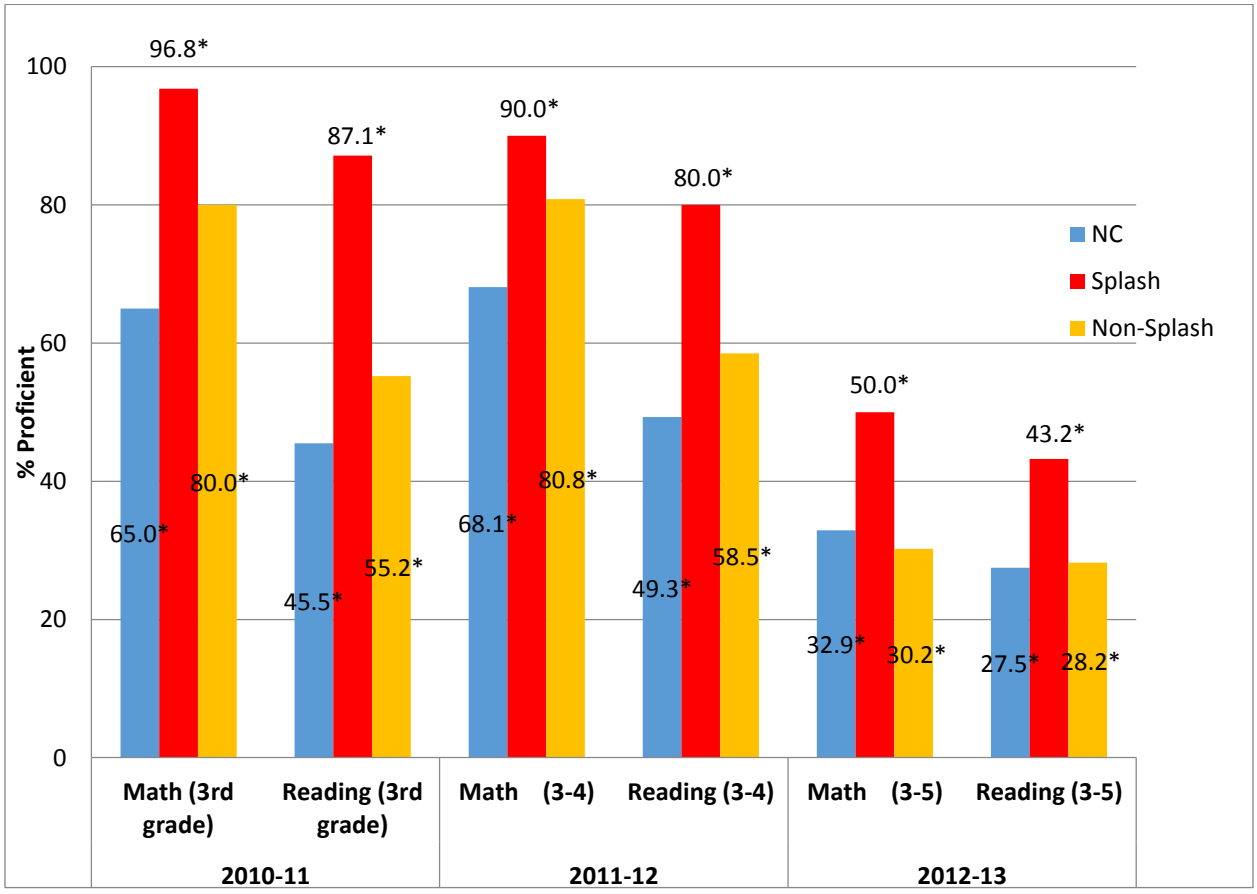


**indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05*

As illustrated in Figure 11, a significantly higher proportion of Splash students eligible for free and reduced price lunch (FRL) demonstrated proficiency on the EOG tests in Reading and Math in 2010-11, 2011-12 and 2012-13, than their FRL eligible peers in the same schools. Since Annual Measureable Objectives targets include proficiency rates for students identified as economically disadvantaged, Splash students again support their schools in reaching both federal and state targets. In addition, a significantly higher proportion of FRL eligible Splash students demonstrated proficiency in Reading and Math in 2010-11, 2011-12 and 2012-13, than their economically disadvantaged peers statewide. There were too few Splash students eligible for free or reduced price lunch with 5th grade Science EOG scores to report results for this test.

Again, as noted above, proficiency rates decrease among all students in 2012-13 due to changes in the End of Grade tests in this year.

Figure 11: Proficiency on EOG Tests, Students Eligible for Free or Reduced Price Lunch, 2010-11-2012-13



**indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05*

Students Identified as Academically or Intellectually Gifted

Splash students identified as academically or intellectually gifted (AIG) perform similarly to their AIG identified peers on Math and Reading EOG tests. In 2010-11, there were too few Splash students identified as AIG to report results for this year. In 2011-12, AIG identified Splash students scored an average of 362.5 on the Math EOG test, while their AIG identified peers in the same school scored an

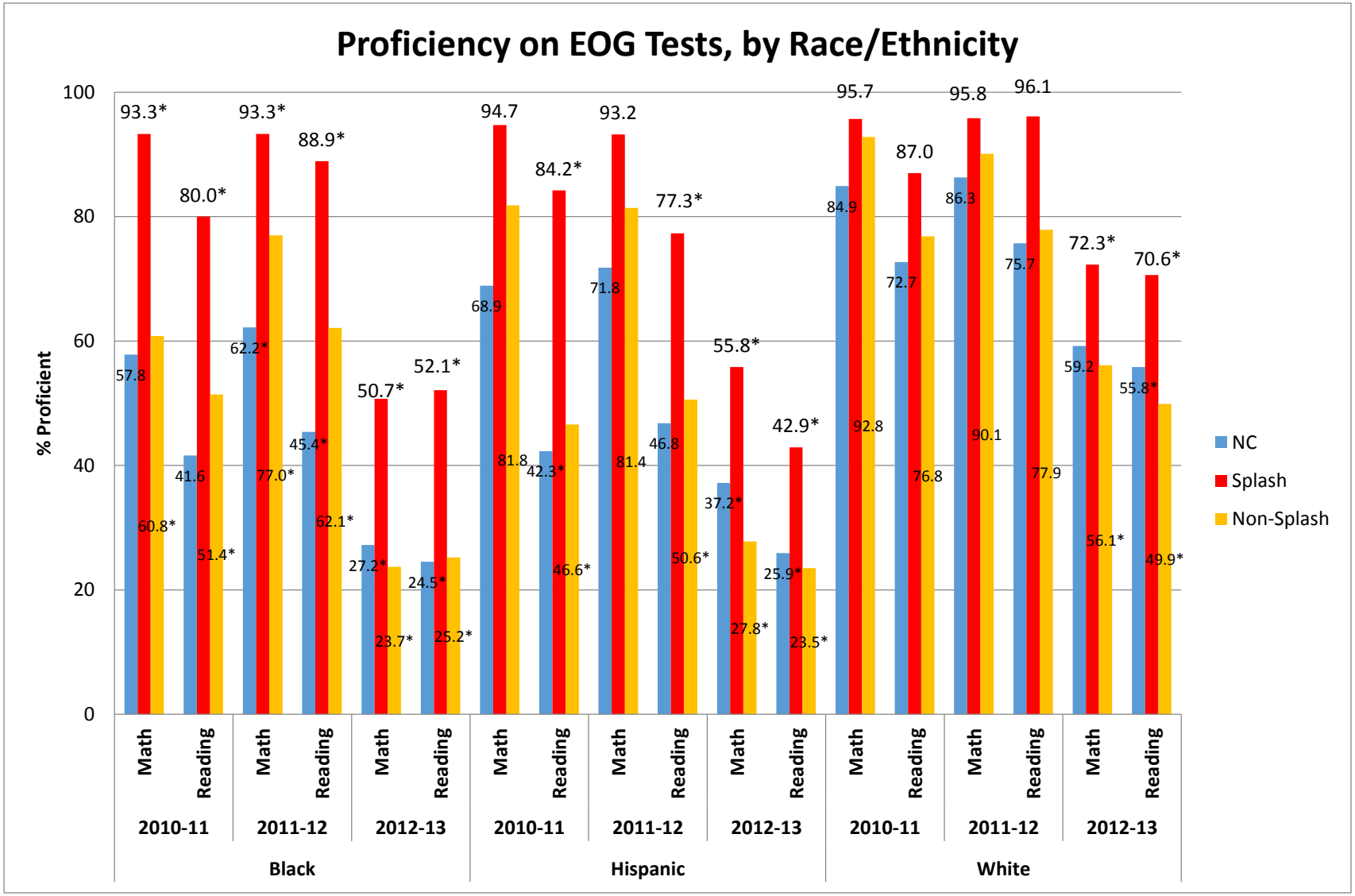
average of 362.2; AIG identified Splash students scored an average of 357.5, while their AIG identified peers scored an average of 357.6. All AIG identified students in Splash schools achieved proficiency on Math and Science EOG tests in 2011-12. In 2012-13, AIG identified Splash students and their peers again performed similarly on EOG tests, with average Math scores of 460.6 as compared to 470.9 for their peers, and average Reading scores of 456.9 as compared to 456.8 for their peers. Proficiency rates were also very similar among AIG students in 2012-13: 97.1% of Splash students achieved proficiency in Math as compared to 97.6% of non-Splash students; 91.2% of Splash students achieved proficiency in Reading as compared to 89.7% of non-Splash students.

Student achievement, by Race/Ethnicity

When Splash student scores are examined by racial and ethnic subgroup, Splash students perform at least as well, and often better than their peers in the same schools and statewide. As seen in Figure 12, among Black students, Splash participants demonstrated significantly higher rates of proficiency than their peers in the same schools on Math and Reading EOG tests in 2010-11, 2011-12, and 2012-13. In addition, Black Splash students demonstrated higher rates of proficiency than Black students statewide on Math and Reading EOG tests in 2011-12 and 2012-13. Among Hispanic students, Splash participants demonstrated higher rates of proficiency on Reading EOG tests in 2010-11, 2011-12, and 2012-13, and on Math EOG tests in 2012-13. As compared to their Hispanic peers statewide, Splash students identified as Hispanic demonstrated higher rates of proficiency on Reading EOG tests in 2010-11 and 2012-13, and on Math EOG tests in 2012-13. Among White students, Splash students achieved similar rates of proficiency on Math and Reading EOG tests in 2010-11 and 2011-12 as compared to their peers in the same schools and their peers statewide. In 2012-13, White Splash students demonstrated significantly higher rates of proficiency on Math EOG tests than their White peers in the same schools and significantly higher rates of proficiency on Reading EOG tests than their White peers in the same schools and statewide. Annual Measurable Objectives (AMO) targets include proficiency rates for students by race and ethnicity, so when Splash students from these subgroups outperform their peers, they again support their schools in reaching both federal and state AMO targets.

Note: While Figure 12 demonstrates that Splash students' proficiency rates are consistently higher than their peers, this report focuses on differences that are statistically significant at the 95% level ($p<=0.05$).

Figure 12: Proficiency on EOG Tests by Race/Ethnicity, 2010-11 through 2012-13

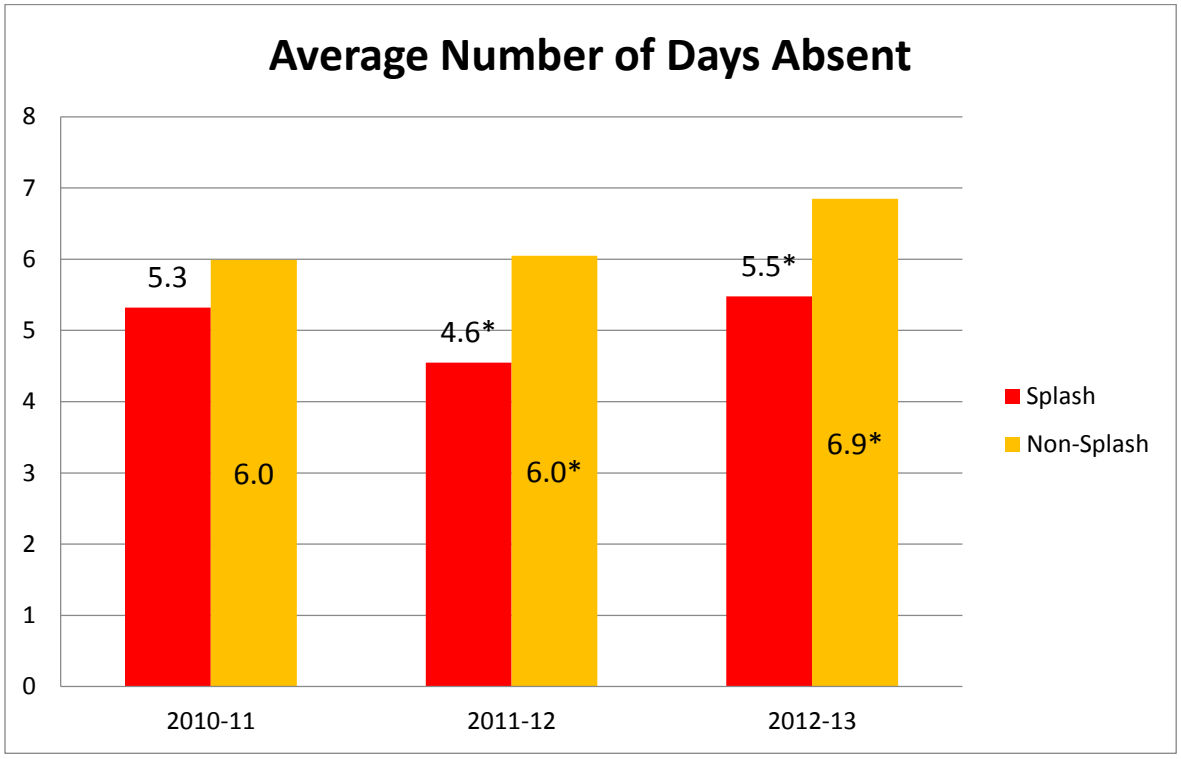


**indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05*

Student Attendance

As illustrated in Figure 13, in 2011-12 and 2012-13, Splash students, on average, were absent fewer days than their peers in the same schools. Since school attendance rates are included in both federal and state Annual Measurable Objectives (AMO) targets, Splash students also support their schools in meeting these targets through their higher rates of attendance.

Figure 13: Average Number of Days Absent, 2010-11 through 2012-13



**indicates a statistically significant difference between the average scores of Splash students and their peers, p<=0.05*

But What About Selection?!?

While descriptive analyses indicate that Splash students are generally outperforming their peers on End of Grade tests, ideally we would like to establish that this improved performance is *caused* by participation in the Splash program. VIF Splash is a voluntary program—the majority of parents decide prior to the beginning of Kindergarten whether to enroll their child in a Splash dual language classroom or in a standard classroom, with a few students entering after the Kindergarten year. Our ability to estimate causal effects is then threatened by *selection bias*--constructing a true comparison group is difficult as we may have individuals who have one set of potential outcomes in the absence of the Splash program in the treatment group (the Splash classrooms) and individuals with a different set of potential outcomes in the absence of treatment in the comparison group. For example, parents may make the choice of classrooms in which they feel that their child would have the best outcomes. It is also possible that there are characteristics of families that choose Splash that would support student success in any classroom. We cannot then simply compare the outcomes of the two groups and determine that any differences are caused by the treatment, in this case, Splash participation.

There are a number of statistical techniques that address selection bias. Since, as discussed on page 5 of this report, Splash participants differ from non-participants on several key observable characteristics, we used propensity score matching to reduce bias. Propensity Score Matching is a statistical technique that uses observed covariates to predict the probability of being in the treatment group (in this case a Splash classroom) and matches treated individuals to untreated individuals with similar probabilities of treatment. After matching, the differences in the outcomes between Splash and non-Splash classrooms are independent of all of the observed characteristics that predict choosing Splash. A wide body of literature supports that propensity score matching techniques offer significant bias reduction, producing estimates that are similar to random assignment studies if the variables that predict treatment are correctly and fully specified (Diaz and Handa, 2005; Cook, Shadish, and Wong, 2008; Henry and Yi, 2009).

We estimated a propensity score for all 3rd grade students, that is the probability of treatment, conditional on several student characteristics. We matched each 3rd grade Splash student to another 3rd grade student at the same school in a non-Splash classroom with the most similar propensity score, limiting the distance between matches to ¼ of a standard deviation of the propensity score. If a student with a similar score could not be found, the student was not included in the sample, rather than matching with a dissimilar student. Table 3 displays all of the demographic characteristics that were significantly different between the two samples, before and after matching. As demonstrated, the two groups are much more similar on these observed characteristics after matching.

Table 3: Splash Students Compared to Non-Splash Students, Before and After Matching

	Before Matching		After Matching	
	Splash Students N=317	**Non-Splash Students in the Same Schools N=1330**	**Splash Students N=289**	**Non-Splash Students in the Same Schools N=289**
Gender				
Male	41.01%	52.48%	41.18%	41.18%
Female	58.99%	52.48%	58.82%	58.82%
Hispanic	29.65%	23.38%	26.99%	26.64%
Free/Reduced Lunch	49.53%	65.71%	51.90%	53.29%
Exceptional Children	2.21%	11.73%	2.42%	2.08%
AIG	6.94%	2.41%	5.54%	5.88%
Overage for Grade	2.84%	18.05%	2.77%	2.08%

After matching, multiple regression analysis was conducted, including covariates that are likely to be related to being in a Splash classroom as well as to EOG test scores, as well as the propensity score in the models to further reduce bias (Glazerman, Levy, and Myer, 2003). Control variables included days absent, race and ethnicity, Limited English Proficiency, Academically/Intellectually Gifted and Exceptionality identification, Free/Reduced Lunch Eligibility, Underage for grade, Overage for grade, and Mobility (whether a student moved in the prior or current school year). Interaction terms were also included to determine if students from particular subgroups have different outcomes in Splash

classrooms than students overall. Finally, school fixed effects were used to limit the comparison to students in the same schools, to make the groups as comparable as possible.

As seen in Table 4, on average, a student in a Splash classroom can expect to score more than 22% of a standard deviation higher on 3rd grade Math EOG tests and nearly 25% of a standard deviation higher on 3rd grade Reading EOG tests than a student in a comparison classroom. It may be useful to think of this impact in terms of days of learning. To put effect sizes in this context, we calculate the average elementary school student's gain on the EOG test for the school year, and translate this into a gain per day of instruction, based on a 180 day school year. On average, Splash students gain the equivalent of 108 more days of reading instruction than a similar student in a non-Splash classroom.[1] We could not estimate days equivalency for the Math EOG test, as the average student did not see any gain on this test during the 12-13 school year, presumably due to a drastic change in test format. We find no additional impact on test score performance for Splash students identified as Hispanic, Black, or Limited English Proficient. Perhaps the most interesting finding is the effect on the Math EOG test scores of students eligible for free and reduced price lunch. We find that on average, an economically disadvantaged student in a Splash classroom can expect to score an additional 32% of a standard deviation higher on the 3rd grade Math EOG test than a free or reduced price lunch eligible student in a comparison classroom. For FRL eligible students in Splash classrooms, this means an average increase of more than ½ of a standard deviation on the Math EOG test, which is enough to compensate for the estimated gap in test scores between economically disadvantaged students and their higher income peers (included in Table 4).

Table 4: Effect of Splash on 3rd grade EOG Test Scores, using a Matched Sample

	Effect on Math EOG Score (Standardized Score) N=571	**p-value**	**Effect on Reading EOG Score (Standardized Score) N=571**	**p-value**
Splash	**0.224*** (0.100)	p=0.026	**0.248*** (0.099)	p=0.013
Hispanic*Splash	-0.048 (0.198)	P=0.810	-0.205 (0.222)	p=0.357
Black*Splash	0.016 (0.146)	p=0.915	0.003 (0.156)	p=0.983
Limited English Proficient*Splash	0.058 (0.226)	p-0.797	0.215 (0.249)	p=0.389
Free/Reduced Lunch*Splash	**0.315*** (0.135)	p=0.020	0.234 (0.153)	p=0.127
Free/Reduced Lunch	**-0.541*** (0.127)	p=0.000	**-0.344*** (0.134)	p=0.011

*Notes: Scores are standardized with a mean of 0 and a standard deviation of 1; Standard Errors, clustered at the classroom level are in parentheses; * indicates statistical significance, p<=0.05*

[1] This calculation makes the assumption that Splash and non-Splash prior scores would be similar. Days Equivalency Equation=(((Effect on EOG Score x Standard Deviation/(Average Yearly Gain for all Elementary Students))) X 180 days of instruction. For more explanation, see Patterson, K.M. and Bastian, K.B. (2014). *UNC Teacher Quality Research: Teacher Portals Effectiveness Report*. Chapel Hill, NC: Education Policy Initiative at Carolina.

Discussion

Overall, we find evidence that Splash is successfully accomplishing the goal of having students in the program at or above grade level in terms of academic achievement. Splash students have higher EOG test scores, on average, than non-Splash students in the same schools and their peers across North Carolina. Additionally, Splash students demonstrate higher rates of proficiency on EOG tests, supporting their schools in meeting Annual Measurable Objective (AMO) targets. Controlling for selection into Splash based on observable characteristics, Splash students, on average, have considerably higher 3rd grade Math and Reading EOG test scores. Although they are underrepresented in the Splash program, we find evidence that economically disadvantaged students gain additional benefit from Splash classrooms on Math EOG tests.

There are some key differences in demographics between Splash and non-Splash students which suggests that there may be issues with recruitment at the school level. Males, students eligible for free and reduced price lunch, and students identified as Exceptional Children are underrepresented in the Splash program, while students identified as academically and intellectually gifted are overrepresented as compared to non-Splash classrooms. This finding suggests that attention should be paid to recruitment and information dissemination practices to ensure that information about the Splash program is reaching all student subgroups.

Limitations

While we have employed advanced statistical techniques to control for selection bias, a major limitation of this study is that because Splash is a voluntary program, there may be unobservable characteristics of families that choose Splash over traditional classrooms that affect student test score performance. While we have controlled for all observed characteristics that predict Splash participation, it is likely that attitudes toward education and other unobserved characteristics affect whether a family chooses Splash and these characteristics likely affect student outcomes. Another limitation is that we have no data on student performance prior to the end of 3rd grade when students have been participating in Splash for nearly 4 school years. Ideally, we would have a measure of student ability or achievement pre-Splash, in order to allow for a better comparison of Splash and non-Splash students.

Recommendations

Based on the findings from this study, the author recommends the following next steps:

1. Discuss recruitment strategies with school administrators of participating schools to ensure that all potential students are receiving information about the Splash program.

2. Analyze 2013-14 student data, when available. A much larger Splash participant student sample will be included in the 2013-14 school year. Analyzing this data will give additional statistical power to detect effects of the Splash program as well as adding an additional year of data to allow for the examination of Splash student performance over time.

a. Compare one-way and two-way Splash programs to determine if these two models impact student performance differently.

3. Further analyze the impact of Splash on economically disadvantaged students. Initial findings support that Splash may be useful in narrowing the gap in test scores between students eligible for free or reduced price lunch and their higher income peers. Several mechanisms are posited for this effect, including peer effects, class size, and teacher credentials. Additional research into this population would be informative for future recruitment strategies.

Acknowledgements

This report would not have been possible without several valuable contributions. I would like to recognize Alisa Chapman with the University of North Carolina General Administration as well as Julie T. Marks, with the Education Policy Initiative at Carolina for generously providing much of the data on which this report is based. I also wish to thank Elizabeth D'Amico with the Education Policy Initiative at Carolina for assisting with the process of requesting permission for data use. I would also like to acknowledge David Young, Vicky Kim, and the staff at VIF international education for providing data to identify Splash classrooms as well as valuable feedback throughout the evaluation process. Last but certainly not least, I would like to thank Gary T. Henry of Vanderbilt University for suggesting the research partnership with VIF.

About the Author

Kristina M. Patterson is a Graduate Research Fellow with the Education Policy Initiative at Carolina (EPIC). Through her work with EPIC, she has contributed to a range of research projects, including several quantitative studies on teacher quality in partnership with the University of North Carolina General Administration, as well as quantitative and qualitative analyses of several Race to the Top initiatives in North Carolina public schools. In addition, she has conducted a range of quantitative analyses for the evaluation of programs in Durham Public Schools, such as Citizen Schools and G.R.E.A.T.

Please direct all questions concerning this report to Kristina M. Patterson at kpatt@unc.edu.

References

Cook, T.D., Shadish, W.R. and Wong, V.C. (2008). Three Conditions Under which Experiments and Observational Studies Produce Comparable Causal Estimates: New findings from Within-Study Comparisons. *Journal of Policy Analysis and Management, 27*: 724-750.

Diaz, J.J. and Handa, S. (2005). An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator: Evidence from Mexico's PROGRESA program. *Journal of Human Resources, 41*: 319-345.

Glazerman, S., Levy, D.M., and Myers, D. (2003). Nonexperimental versus Experimental Estimates of Earnings Impacts. *Annals of the American Academy of Political and Social Science, 589*: 63-93

Henry, G. T., and Yi, P. Design Matters: A Within - Study Assessment of Propensity Score Matching Designs. Paper presented at the meeting of the Association for Public Policy Analysis and Management, Washington, D.C., Nov. 7, 2009.

North Carolina Department of Public Instruction. (2011). The North Carolina State Testing Results, 2010-11. Available online at: http://www.ncpublicschools.org/accountability/testing/reports/green/archive

North Carolina Department of Public Instruction. (2012). The North Carolina State Testing Results, 2011-12. Available online at: http://www.ncpublicschools.org/accountability/testing/reports/green/archive

North Carolina Department of Public Instruction. (2013). The North Carolina State Testing Results, 2012-13. Available online at http://www.ncpublicschools.org/accountability/testing/reports/archive

Patterson, K.M. and Bastian, K.B. (2014). *UNC Teacher Quality Research: Teacher Portals Effectiveness Report*. Chapel Hill, NC: Education Policy Initiative at Carolina. Available online at: http://publicpolicy.unc.edu/files/2014/06/Teacher-Portals-Effectiveness-Report.pdf

Rosenbaum, P. R. and Rubin, D.B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician 39*: 33–38.