The Impact of BrainPOP on State Assessment Results

A study of the effectiveness of BrainPOP in grades 3-8

This report was prepared and published by BrainPOP LLC ("BrainPOP").

To provide feedback on this report, contact feedback@brainpop.com.

Abstract

This study compared the academic performance of new BrainPOP® subscribers to non-subscribers using the results of statewide tests taken at the end of the 2015-2016 school year. The results qualify as Moderate ESSA Evidence, showing that schools with a BrainPOP subscription had a greater increase in standardized state test scores than a matched control group in all three subject tests: Math, ELA, and Science. The effect was always positive, always statistically significant at the $p<0.10$ level, and verified in five states. To further validate the results, two additional correlational analyses that qualify as Promising ESSA Evidence were also performed. These analyses found generally positive results that were often statistically significant. The strongest effects were in grade 3-6 and in Math and Science.

Introduction

BrainPOP is an educational platform that offers digital content across the curriculum including animated movies, coding projects, student creation and reflection tools, learning games, and interactive quizzes. BrainPOP also offers various forms of support for that content such as customizable and playful assessments, lesson plans, and professional development opportunities. The BrainPOP suite of creation and reflection tools offers students a range of options for conveying comprehension, ensuring that all children—regardless of learning style—have a way to express their understanding of the content. BrainPOP has continued to hone and expand its content and features over the years, with the goal of providing the broadest range of students agency to explore and "show what they know" in creative and personalized ways.

The BrainPOP philosophy is that there is no single "right" way to use the platform. Similarly BrainPOP provides teachers with the most high quality non-prescriptive resources possible, shaped by relationships with educators and intended for use in ways that educators feel are most effective for their particular students. Through their BrainPOP subscription, all teachers have access to multiple options for teaching and assessment, and every teacher chooses a different path through these resources.

This efficacy study takes a broad perspective on the use of BrainPOP in schools. Considering the non-prescriptive nature of BrainPOP's relationship to teaching practices, it's difficult to consider what implementation with fidelity may look like. For the purposes of this research, we considered the broadest use case—simply being a BrainPOP subscriber—to be the most inclusive intervention category that best accommodates the multitude of use cases that occur with BrainPOP. Research that considers which use cases of BrainPOP are most effective at fostering student achievement will be left to further studies; this study determined whether the use of BrainPOP in some form generally leads to higher student performance.

The approach to efficacy used in this paper also allows a determination of product value at a large scale. Studies of efficacy or effectiveness frequently focus on a single district or state. In contrast, in this one analysis we were able to extend our methodology across five states, each with a different achievement test. This was possible for a few reasons. First, since BrainPOP's presence in schools is so widespread, a focused pilot or study was not necessary to achieve a sample size large enough to conduct a study. Secondly, newer data science techniques that allow for careful selection of matched control groups made it possible to satisfy the new ESSA evidence standards for Moderate evidence with publicly available data. As a result, we were able to conduct the equivalent of five studies in five states by cross applying one technique with relatively little additional effort.

Methods

Data

For this study, we included data from the 2015-2016 school year. A school was considered a BrainPOP subscriber (intervention group) if it had an active subscription to BrainPOP for the entire school year (September 2015 to June 2016). The non-subscriber group (control group) included schools that did not subscribe to BrainPOP, as well as those that had a subscription that either started or ended mid-way through the school year. BrainPOP offers multiple products, but only a subscription to the flagship BrainPOP product was used to segment schools into the intervention and control groups. It is worth noting that many of these schools had a "Combo" subscription, which included BrainPOP Jr. (K-3), BrainPOP Español, and BrainPOP Français. This analysis was also limited to K-8 public schools due to ease of access to public school test scores; private schools (both subscribers and non-subscribers) were excluded from the analysis.

We chose five states for the analysis that best fit a mix of the following criteria: available and easily accessible public test score data, public test score data that used raw numbers rather than percentiles for schools, states with a significant BrainPOP subscriber base, and states with a relatively large number of schools. The last two criteria were intended to ensure relatively large sample sizes in both the intervention and control groups to best aid statistical testing. These criteria led to the selection of the following five states: California, Colorado, Florida, New York, and Texas.

For California, data came from the Smarter Balanced test. Math, ELA, and Science scores were only available for grades 5 and 8 due to California's lack of public data release for other grades[1].

For Colorado, data came from the PARCC (Partnership for Assessment of Readiness for College and Careers) test. Math and ELA scores were available for grades 3 through 8; Science scores were only available for grades 5 and 8[2].

For Florida, data came from the FSA (Florida Standards Assessment) test. Math and ELA scores were available for grades 3 through 8; Science scores were only available for grades 5 and 8[3].

---

[1] https://caaspp.cde.ca.gov
[2] https://www.cde.state.co.us/assessment
[3] http://www.fldoe.org/accountability/assessments/k-12-student-assessment/results/

For New York, data came from the NYSA (New York State Assessment). Math and ELA scores were available for grades 3 through 8; Science scores were only available for grades 4 and 8[4].

For Texas, data came from the STAAR (State of Texas Assessment of Academic Readiness). Math and ELA scores were available for grades 3 through 8; Science scores were only available for grades 5 and 8[5].

We pulled demographic information about schools from the NCES database[6]. Information in the database included school location, grade level of the school, number of students, number of teachers, and socioeconomic indicators.

Statistics

We used three different methodologies to compare our intervention group to the control group. All statistics were performed in R[7]. In all cases, Cohen's D was used to calculate the effect size, and an unpaired Student's t-test was used to test for significance using the default t.test function in R. Since many of the designs had unequal sample sizes, this function often defaulted to the Welch's t-test, correcting the degrees of freedom for unequal sample sizes and variance. Methods A and B qualify as a correlational study and pass the Promising level of evidence for ESSA. Method C qualifies as a quasi-experimental study, and passes the Moderate level of evidence for ESSA.[8]

The first method (Method A in the Detailed Results section) corrected for the effect of a single socioeconomic factor, namely percentage of students with free or reduced lunch, between the experimental and control group. This method counted as a correlational study that used analytic methods to control for intervention and control group differences. We found that free or reduced lunch percentages were correlated with standardized test scores in some states but not others (results not shown in this report[9]). For example, New York had a strong negative correlation between free and reduced lunch percentage and test scores, while Florida had a weak negative

---

[4] https://data.nysed.gov/downloads.php
[5] https://tea.texas.gov/student.assessment/staar/aggregate/
[6] https://nces.ed.gov/ccd/schoolsearch/
[7] https://www.r-project.org/
[8] Newman, D., Jaciw A, and Lazarev V. 2018. Guidelines for Conducting and Reporting EdTech Impact Research in U.S.K-12 Schools. Whitepaper from Empirical Education and Education Technology Industry Network of SIIA
[9] Although a positive correlation between socio-economic status and test scores is well-known, the magnitude of the effect does vary from analysis to analysis, with each analysis often focused on a different state. We are not aware of a single comprehensive study that explicitly compares states, but our findings seem consistent with the differing correlation coefficients found in the literature.

correlation, and California had no correlation. Other indicators had similar discontinuities, such as Title I funding showing a strong negative correlation in New York, no general correlation in Florida, and a weak negative correlation in California. In general, free and reduced lunch percentage seemed to have the most consistent and strongest correlation with test scores across states, and so we chose it as the single variable of focus in this method for simplicity, and left corrections that accounted for multiple factors to the next two more complex methods. We used a linear regression to account for the interaction between free and reduced lunch percentage and test scores, deriving a different correction factor for each grade and state. Once corrected, the intervention and control are compared with an unpaired Student's t-test.

The second method (Method B) offered an even more conservative estimate by matching multiple demographic factors between the control and intervention group. This is a matched comparison methodology. Specifically, the populations were matched on these factors: school location (rural, urban, suburban), charter schools, Title 1 funding, percent free or reduced lunch, number of students, and number of teachers. We did this using the Frontier Matching Method using the MatchingFrontier R package[10].

In short, this method trims the control and intervention population until their distribution on all the included demographic variables is matching, or in other words, it removes outliers from both the control and intervention groups (for resulting sample size used in each statistical test with this method, see Appendix 1). This process resulted in smaller samples sizes than Method A, but in populations who are matched in all demographic characteristics deemed important. We chose parameters such that the control and intervention groups were less than 0.25 standard deviations away from each other on all relevant characteristics. As a result of this method, for every BrainPOP school, there were potentially multiple non-BrainPOP schools matched based on the similarity of their characteristics. For more details on the Frontier methodology, see the documentation[11]. Once this matching process was complete, we compared the two resulting populations with an unpaired Student's t-test.

The third method (Method C) offered a different perspective, focusing solely on schools in their first year of using BrainPOP. The sample sizes were much smaller in this analysis because we limited our approach to schools that subscribed to BrainPOP for the first time between June 2015 and March 2016. We used the 2015 state test as a pre-test and the 2016 state test as a post-test. This approach, which qualified as quasi-experimental, identified the school which was demographically most similar to each school in the intervention group using a nearest neighbor method as described in the MatchIt R package[12], and put that school in the control group. For the

[10] https://cran.r-project.org/web/packages/MatchingFrontier/vignettes/Using_MatchingFrontier.pdf
[11] https://projects.iq.harvard.edu/frontier/home
[12] https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf

demographic profile of the resulting experimental and control groups, see Appendix 2. The sample sizes of the intervention and control group were equal in this method (see Appendix 1 for the sample size used in each grade and state). Specifically, the populations were matched on these factors for the grade-level analysis: urban school location, Title 1 funding, percent free or reduced lunch, number of students, and pretest score (2015 data). For the school-level analysis, schools were matched on the above parameters plus one additional parameter: school grade level (elementary vs. middle school). Then we compared the two matched groups with an unpaired Student's t-test.

To look across all five states we made an additional calculation for Method C. All states directly reported grade-specific test scores—we used those scores to calculate a school-wide average score for schools in the control and intervention groups. We calculated the difference from mean for each grade, and then averaged over all available grades to get a school-wide average difference from mean. We used this school-wide average difference from mean to calculate both the 2015 and 2016 scores. Thus, the control group used for the school aggregated results was a different set of schools than those used in each of the grade-level results. The control group was defined by the same algorithm; however, by using a school-wide average distance from mean as the pre-test factor in the matching process, the algorithm may find a different set of matched schools than those found in the grade-specific results. Thus, the effect sizes reported in the school-aggregated results can potentially show different patterns from what is contained in each individual grade-level result for Method C, even though the same algorithm was used in both methods.

Results

School-aggregated Results

We will start by discussing the results that correspond to proving Moderate ESSA evidence. These results are for Method C using the school-aggregated test scores. For this methodology, we found an average effect size of 38.4 in ELA, 44.8 in Math, and 46.4 in Science across five states, indicating that BrainPOP subscribers perform better than a demographically similar set of non-subscribers.

Figure 1 illustrates the change from pretest to posttest between the intervention group and matched control group. The intervention group of BrainPOP subscribers shows a modest increase in performance, while the matched control group shows a significant decrease in performance. Results in Figure 1 correspond to a school's mean ELA, Math, and Science scores, averaged across the five states in the analysis.

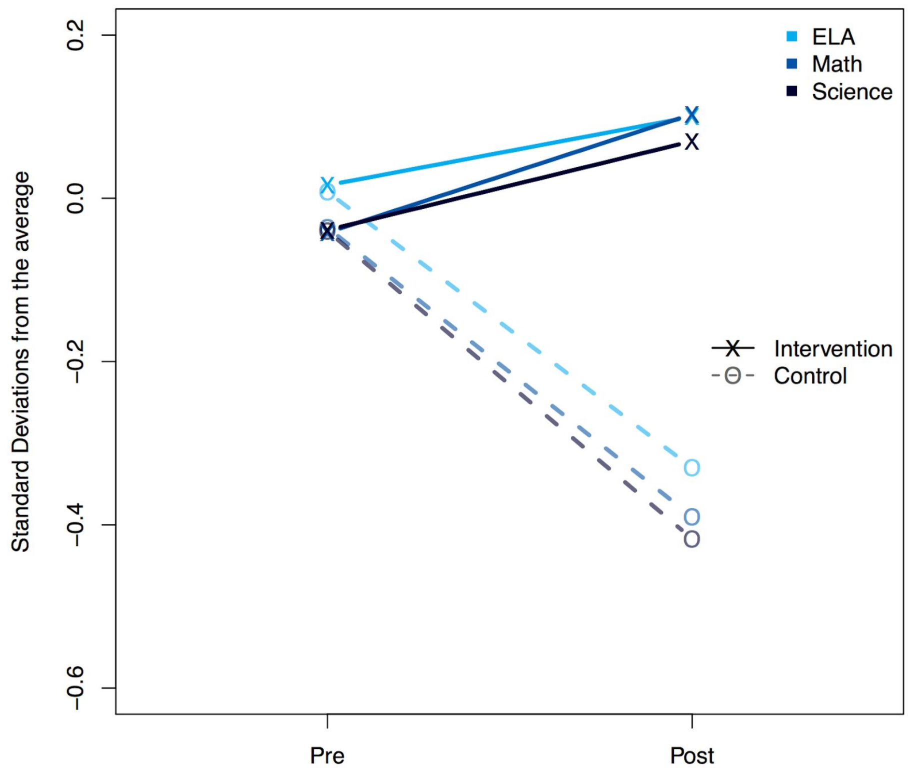

Figure 1 *- Change from pre to post averaged across all five states, measured in units of standard deviation from the mean and then averaged across all available grade levels. The solid, dark lines correspond to the intervention and the dashed transparent lines are for the control group. Both groups were matched on their pretest scores.*

If we look at each state individually, there is a positive effect for each state and subject (Table 1). Additionally, in all cases, the effect is significant at the $p<0.10$ level with 66% significance at the $p<0.05$ level. The effect sizes are also relatively high, which is due in part to the decrease in scores of the control group in each subject test.

	ELA	**Math**	**Science**
California	0.27 **	0.28 **	0.26 **
Colorado	0.50 *	0.73 **	0.65 *
Florida	0.55 *	0.55 *	0.55 **
New York	0.32 *	0.39 **	0.60 ***
Texas	0.28 ***	0.29 ***	0.26 **

Table 1 *- Effect sizes for each state and subject test, based on school-wide scores (difference from mean averaged over all available grades). Significance values are denoted as * p<0.10, ** p<0.05, *** p<0.01*

Overall, using Method C we found a positive effect of having a BrainPOP subscription and that the effect is often statistically significant, which satisfies the ESSA Evidence Standard for Moderate evidence. Results for Method A and B were only evaluated at the grade level and not the school-wide level, and are included in the next section.

Detailed Results by Grade

This section details the results for each individual grade. Each state is also analyzed separately in the study. All three methods produced generally similar results: BrainPOP subscribers had higher test scores than non-subscribers, which confirms the school-averaged results (see Tables 2-6). The magnitude and significance of the effect varied from method to method, but in all five states, a positive effect resulted at every grade level for at least one method, and the effect was often statistically significant. The effect was also typically stronger in grades 3-6, and in Science and Math. In several of the states, grades 7 and especially 8 tended to show non-significant patterns, but even in these cases a positive effect of a BrainPOP subscription was still seen—just not one that was large enough to warrant significance. Effect sizes tended to be larger in Method C but with a lower level of statistical significance due to the smaller sample sizes necessitated by that methodology.

Note that one important characteristic to consider is that of the BrainPOP subscribers included in Method C, there was a high rate of schools that qualify for Title 1 funding. In specific, the BrainPOP schools were 15% - 65% more likely to qualify for Title 1 status than the average school, depending on the state (see Appendix 2 for a full list of the demographic characteristics of schools included in Method C). Thus, BrainPOP subscribers tend to be among a demographic group of schools that tend to be more underperforming on average[13], yet they seem to be showing improvements in their average test score after their first year with BrainPOP.

It is worth noting that as correlational methods, Methods A and B make a very different point than Method C. Both of the former methods included many long term BrainPOP subscribers, or schools that had subscribed to BrainPOP for several years before the 2016 scores were taken. BrainPOP has existed as a subscription service since 1999 and many schools included in Method A and B subscribed to BrainPOP for more than five, and in some cases even ten, years. The 2015-2016 school year was not their first year of using BrainPOP and it doesn't make sense to consider them as new subscribers undergoing an intervention, and therefore to use a pre-post methodology. Thus the purpose of these correlational methods was mainly to determine whether long-term subscribers who have had many years of experience with BrainPOP tend to score

[13] The correlation between socio-economic status and test scores is reported throughout the literature, but for one example: news.stanford.edu/2016/04/29/local-education-inequities-across-u-s-revealed-new-stanford-data-set/

higher on standardized tests than those schools which do not subscribe to BrainPOP. Although only considered Promising evidence by the ESSA standards, these methods still prove an important point and offer a complementary corroboration for the positive results found by the Method C quasi-experimental intervention.

California

	ELA			Math			Science		
	A	B	C	A	B	C	A	B	C
grade 5	0.07**	0.20***	0.27**	0.04	0.18***	0.27**	0.11***	0.20***	0.28**
grade 8	0.02	0.12**	0.04	0.14***	0.10**	-0.01	0.11**	0.23***	0.19

Table 2 *- Effect sizes for each grade and subject test for the California Smarter Balanced test. Significance values are denoted as * p<0.10, ** p<0.05, *** p<0.01*

California is exceptional because at the time we pulled the data, test scores were only available for two grades. From this data set, every effect size except one (grade 8 math, Method C) was positive, and 72% of the effect sizes are significant at the p<0.05 level, showing an overall positive and significant effect of the intervention (Table 2).

Science had the most statistically significant results, and the largest effect sizes**.** Also, grade 5 showed larger effect sizes and more statistically significant results than grade 8.

Colorado

	ELA			Math			Science		
	A	B	C	A	B	C	A	B	C
grade 3	0.08***	0.02	0.73**	0.14***	0.19**	0.79**			
grade 4	0.10***	0.24***	0.68**	0.13***	0.15	0.86**			
grade 5	0.10***	0.33***	0.03	0.17***	0.23*	0.64*	0.07**	-0.02	0.15
grade 6	0.07**	0.11	0.08	0.07**	0.12	0.05			
grade 7	0.11**	0.15	0.15	0.10***	0.09	0.35			
grade 8	0.07	0.11	0.13	0.09***	0.21*	0.19	0.02	-0.04	-0.01

Table 3 *- Effect sizes for each grade and subject test for the Colorado PARCC test. Significance values are denoted as * p<0.10, ** p<0.05, *** p<0.01*

Table 3 shows positive effect sizes in Colorado across all grades for Math and ELA**.** The results revealed some differences by methodology: Method A showed significant results across grades and test. Methods B and C showed generally positive results, but only some statistical significance across grades and subjects. Colorado had the lowest sample sizes (Appendix 1), which can in part explain the higher variability seen in the results. Grades 3-5 showed the strongest effects overall across methods; Method C showed particularly large effect sizes in these grades for ELA and Math.

Florida

	ELA			Math			Science		
	A	B	C	A	B	C	A	B	C
grade 3	0.22***	0.27***	0.52*	0.10**	0.12**	0.67**			
grade 4	0.17***	0.26***	0.59*	0.13***	0.16***	0.57*			
grade 5	0.13***	0.14***	0.40	0.10**	0.04	0.45	0.23***	0.26***	0.45*
grade 6	0.22***	0.25***	0.11	0.12***	0.23***	0.27			
grade 7	0.17***	0.27***	0.31	0.17***	0.11	0.19			
grade 8	0.08	0.19**	0.01	0.09	0.12	0.45	0.18***	0.12	0.16

Table 4 *- Effect sizes for each grade and subject test for the Florida FSA test. Significance values are denoted as * p<0.10, ** p<0.05, *** p<0.01*

Florida had positive results across grade and methodology as seen in Table 4. All effect sizes were positive, with 55% of the tests showing statistical significance at the p=0.05 level**.** Grade 8 is the exception, showing lower effect sizes and statistical significance than the other grades. The results in Florida were equally positive across the three different subject tests.

Method C's results were generally positive but not statistically significant. This may be due to low sample sizes. The effect sizes themselves were similar to those found in Methods A and B, indicating that a larger sample size might have resulted in statistical significance.

New York

	ELA			Math			Science		
	A	B	C	A	B	C	A	B	C
grade 3	0.08*	0.23***	0.42**	0.09**	0.16***	0.40**			
grade 4	0.17**	0.25***	0.35*	0.14***	0.20***	0.45**	0.13***	0.21***	0.40**
grade 5	0.06	0.22***	0.21	0.13***	0.20***	0.34			
grade 6	0.10*	0.27***	0.25	0.17***	0.24***	0.33			
grade 7	0.07	0.06	0.06	0.12**	0.13**	0.26			
grade 8	0.07	0.13*	0.04	0.07	0.12*	0.78**	0.03	0.10	0.56

Table 5 *- Effect sizes for each grade and subject test for the New York NYSA test. Significance values are denoted as * p<0.10, ** p<0.05, *** p<0.01*

The New York results are overwhelmingly positive across grade and methodology, with 52% of the tests resulting in statistical significance at the p=0.05 level (Table 5). Most non-significant results were in grades 7 and 8. The effect sizes were positive in every grade. Method C had less statistical significance than the other methods due to a smaller sample size, but often had larger effect sizes. The effect sizes tended to be slightly larger in Math and Science than in ELA.

Texas

	ELA			Math			Science		
	A	B	C	A	B	C	A	B	C
grade 3	0.01	0.01	0.27**	0.02	0.01	0.28**			
grade 4	0.27***	0.25***	0.23*	0.25***	0.27***	0.25**			
grade 5	0.27***	0.24***	0.03	0.30***	0.28***	0.26*	0.30***	0.28***	0.32**
grade 6	0.13***	0.07	0.23	0.19***	0.13***	0.15			
grade 7	0.12**	0.00	0.31*	0.13***	0.03	0.04			
grade 8	0.11**	0.00	-0.01	0.25***	0.18***	0.33*	0.29***	0.19***	0.24

Table 6 *- Effect sizes for each grade and subject test for the Texas STAAR test. Significance values are denoted as * p<0.10, ** p<0.05, *** p<0.01*

The results for Texas were generally promising, with 57% of the tests significant at the p=0.05 level (Table 6). The results for Texas were a little more variable across grades and methodology than the other states. In all cases except one (grade 8 ELA Method C), the effect sizes were positive. Grades 4 and 5 showed consistent results across methodology and subject tests, whereas the other grades tended to show more variability. In every grade and every subject test, at least one of the three methods showed a statistically significant positive effect, confirming an overall positive pattern across grades. Methods A and B tended to show higher effect sizes in Math and Science compared to ELA, whereas all subject tests performed relatively equal in Method C.

Conclusion

Overall, having a BrainPOP subscription produces a positive effect on state test scores across grades and subjects. This is true for both the correlational methods (Methods A and B) that focused on long-term subscribers, and the quasi-experimental method (Method C) that focused on subscribers in their first year of using BrainPOP. The latter methodology qualifies the results as Moderate ESSA Evidence.

This positive effect of a BrainPOP subscription also held across all five states included in the study, indicating a level of replicability to the results. Further analyses can investigate the type of BrainPOP usage that most contribute to increasing student achievement.

Appendix 1

Appendix 1 shows the sample size, or number of total schools, used in the analyses performed for each state and grade, including both the intervention and the control group. The appendix also includes the sample sizes in the school-aggregated results performed for Method C. The sample size shown in the chart accounted for all schools that had test score data available for that grade and any additional selection criteria used for that statistical method. In general, Method A had the largest sample sizes since it included all schools in a state for which test scores existed and simply added a correction for those scores. For Method B, the sample size was the number of schools left in the intervention and control group after the Frontier trimming process. Method B had smaller sample sizes than Method A, but was still quite large, typically including 60-70% of all schools in the state. Method C had much smaller sample sizes due to the strict nature of the selection process for this methodology.

California

	ELA			Math			Science		
State	A	B	C	A	B	C	A	B	C
grade 5	4771	3839	206	4775	3856	208	4783	3867	202
grade 8	2031	1594	80	2042	1600	90	2045	1456	90
Aggregate			218			232			226

Table A1.1 *- The sample size for each method for California.*

Colorado

	ELA			Math			Science		
State	A	B	C	A	B	C	A	B	C
grade 3	934	383	36	941	473	34			
grade 4	935	431	36	935	467	34			
grade 5	926	391	30	930	491	28	927	450	26
grade 6	554	305	18	555	351	18			
grade 7	445	246	12	440	247	12			
grade 8	428	240	12	405	247	12	432	273	10
Aggregate			48			44			30

Table A1.2 *- The sample size for each method for Colorado.*

Florida

	ELA			Math			Science		
State	A	B	C	A	B	C	A	B	C
grade 3	1797	1436	40	1805	1454	38			
grade 4	1783	1433	42	1800	1445	38			
grade 5	1758	1603	42	1772	1453	38	1802	1457	54
grade 6	979	796	32	986	627	34			
grade 7	916	644	34	937	543	32			
grade 8	921	771	34	921	540	32	909	519	32
Aggregate			46			42			56

Table A1.3 *- The sample size for each method for Florida.*

New York

	ELA			Math			Science		
State	A	B	C	A	B	C	A	B	C
grade 3	2149	1696	92	2214	1717	94			
grade 4	2113	1736	92	2184	1701	94	2604	1795	84
grade 5	2091	1667	80	2105	1616	68			
grade 6	1406	1125	40	1478	1159	40			
grade 7	1241	1007	36	1270	1000	36			
grade 8	1182	960	32	1149	958	28	1166	964	22
Aggregate			124			112			102

Table A1.4 *- The sample size for each method for New York.*

Texas

	ELA			Math			Science		
State	A	B	C	A	B	C	A	B	C
grade 3	4234	3346	254	4234	3346	254			
grade 4	4181	3347	250	4181	3284	252			
grade 5	3849	3021	206	3849	3021	204	3849	3021	202
grade 6	2109	1703	124	2109	1703	126			
grade 7	1814	1479	128	1814	1479	110			
grade 8	1863	1524	108	1863	1524	106	1863	1524	106
Aggregate			194			410			292

Table A1.5 *- The sample size for each method for Texas.*

Appendix 2

Appendix 2 shows a demographic comparison between the BrainPOP subscriber and non-subscriber groups for each state for all of the variables used in Method C's matching technique for the school-aggregated results. The variables are shown in the format of Mean ± 0.25 Standard Deviations.

California

	ELA		Math		Science	
	Inter.	Con.	Inter.	Con.	Inter.	Con.
2015 score	2.7 ± 10.3	2.6 ± 10.4	2.9 ± 11.1	3.0 ± 11.0	0.8 ± 9.1	0.9 ± 9.2
% Urban	46.8% ± 12.5%	48.6% ± 12.6%	46.6% ± 12.5%	49.1% ± 12.6%	47.8% ± 12.5%	44.6% ± 12.5%
% Title 1 funding	67.0% ± 10.7%	67.9% ± 11.7%	60.0% ± 12.3%	66.4% ± 11.9%	60.2% ± 12.3%	66.1% ± 11.9%
Average FRL percentage	59.7% ± 7.2%	62.7% ± 7.1%	63.0% ± 6.6%	63.0% ± 6.6%	58.4% ± 7.2%	65.1% ± 6.9%
Average number of students	660 ± 73	612 ± 66	664.7 ± 72.8	637.2 ± 64.0	660.0 ± 71.8	646.9 ± 68.1
% Elementary schools	70.6% ± 11.4%	74.3% ± 11.0%	69.8% ± 11.5%	69.8% ± 11.5%	72.6% ± 11.2%	73.2% ± 11.1%

Table A2.1 *- The demographic distribution for the intervention and control groups in Method C for California.*

Colorado

	ELA		Math		Science	
	Inter.	Con.	Inter.	Con.	Inter.	Con.
2015 score	-4.7 ± 3.1	-4.8 ± 3.0	-3.1 ± 2.6	-3.1 ± 2.6	-8.8 ± 11.8	-7.4 ± 12.0
% Urban	25.0% ± 11.1%	29.2% ± 11.6%	22.7% ± 10.7%	31.8% ± 11.9%	26.7% ± 11.4%	26.7% ± 11.4%
% Title 1 funding	45.8% ± 12.7%	37.5% ± 12.4%	45.5% ± 12.7%	45.5% ± 12.6%	46.7% ± 12.9%	53.3% ± 12.9%
Average FRL percentage	49.4% ± 6.7%	52.9% ± 6.7%	48.4% ± 6.6%	56.0% ± 7.1%	55.7% ± 6.4%	55.7% ± 6.2%
Average number of students	560.0 ± 67.0	543.5 ± 48.0	567.8 ± 69.7	538.1 ± 57.4	627.5 ± 151.8	627.5 ± 75.2
% Elementary schools	75.0% ± 11.1%	62.5% ± 12.4%	72.7% ± 11.4%	63.6% ± 12.3%	73.3% ± 11.4%	80.0% ± 10.4%

Table A2.2 *- The demographic distribution for the intervention and control groups in Method C for Colorado.*

Florida

	ELA		Math		Science	
	Inter.	Con.	Inter.	Con.	Inter.	Con.
2015 score	3.0 ± 2.3	2.7 ± 2.3	-0.7 ± 2.5	-0.6 ± 2.5	-0.5 ± 2.1	-0.7 ± 2.1
% Urban	26.1% ± 11.2%	34.8% ± 12.2%	30.4% ± 11.8%	21.7% ± 10.5%	25.0% ± 11.0%	21.4% ± 10.4%
% Title 1 funding	82.6% ± 9.7%	73.9% ± 11.2%	82.6% ± 9.7%	78.3% ± 10.5%	92.9% ± 6.6%	85.7% ± 8.9%
Average FRL percentage	46.1% ± 4.7%	51.7% ± 5.8%	56.5% ± 4.8%	62.8% ± 4.4%	52.7% ± 3.9%	53.0% ± 4.6%
Average number of students	1056.3 ± 88.6	1069.8 ± 86.6	1036.3 ± 80.3	1078.7 ± 56.8	1085.8 ± 72.3	1055.1 ± 53.6
% Elementary schools	52.2% ± 12.8%	65.2% ± 12.2%	52.2% ± 12.8%	52.2% ± 12.8%	50.0% ± 12.7%	60.7% ± 12.4%

Table A2.3 *- The demographic distribution for the intervention and control groups in Method C for Florida.*

New York

	ELA		Math		Science	
	Inter.	Con.	Inter.	Con.	Inter.	Con.
2015 score	-0.5 ± 3.2	-0.5 ± 3.1	1.4 ± 4.1	1.7 ± 4.0	0.3 ± 1.7	0.2 ± 1.6
% Urban	62.9% ± 12.2%	62.9% ± 12.2%	62.5% ± 12.2%	50.1% ± 12.6%	62.8% ± 12.2%	58.8% ± 12.4%
% Title 1 funding	79.0% ± 10.3%	72.6% ± 11.2%	76.8% ± 10.7%	80.0% ± 10.1%	72.6% ± 11.3%	78.4% ± 10.4%
Average FRL percentage	61.2% ± 7.1%	60.4% ± 6.6%	59.7% ± 7.2%	59.2% ± 6.6%	57.9% ± 7.4%	43.1% ± 12.5%
Average number of students	516.6 ± 66.8	540.9 ± 67.5	516.6 ± 66.9	503.8 ± 60.5	496.2 ± 63.3	506.6 ± 60.8
% Elementary schools	75.8% ± 10.8%	64.5% ± 12.1%	75.0% ± 10.9%	61.8% ± 12.3%	52.9% ± 12.6%	43.1% ± 12.5%

Table A2.4 *- The demographic distribution for the intervention and control groups in Method C for New York.*

Texas

	ELA		Math		Science	
	Inter.	Con.	Inter.	Con.	Inter.	Con.
2015 score	4.5 ± 10.1	4.7 ± 10.1	0.4 ± 10.2	0.4 ± 10.2	-2.0 ± 44.8	-3.1 ± 45.0
% Urban	28.4% ± 11.3%	32.0% ± 11.7%	28.3% ± 11.3%	31.2% ± 11.6%	25.3% ± 10.9%	28.1% ± 11.3%
% Title 1 funding	85.8% ± 8.8%	91.4% ± 7.0%	85.9% ± 8.7%	83.9% ± 9.2%	86.3% ± 8.6%	89.0% ± 7.8%
Average FRL percentage	63.0% ± 5.9%	51.3% ± 12.5%	62.3% ± 5.9%	63.5% ± 5.6%	60.7% ± 5.4%	62.1% ± 5.1%
Average number of students	543.1 ± 71.4	543.1 ± 71.4	547.5 ± 70.3	555.9 ± 72.4	509.9 ± 70.4	512.5 ± 71.9
% Elementary schools	51.3% ± 12.5%	51.3% ± 12.5%	55.6% ± 12.5%	47.3% ± 12.5%	56.8% ± 12.4%	60.3% ± 12.3%

Table A2.5 *- The demographic distribution for the intervention and control groups in Method C for Texas.*