



BENCHMARK ADVANCE ©2022
& BENCHMARK ADELANTE ©2023

ESSA EVIDENCE



B E N C H M A R K E D U C A T I O N C O M P A N Y

Executive Summary

This report provides Every Student Succeeds Act (ESSA) evidence levels, as well as details and comments, on ten different studies where *Benchmark Advance* and/or *Benchmark Adelante* were used by students as their ELA core program. The studies have occurred, thus far, in Arizona, California, Colorado, Delaware, Florida, Michigan, Minnesota, and North Carolina. California is where *Benchmark Advance* and *Benchmark Adelante* are used by the largest number of districts due to the formal adoption process. The Florida adoption also created a group of districts using *Benchmark Advance* ©2022 and the state test results gave us evidence the program is working. Tables 1 and 2 provide the title of the study, a short title used in the rest of the report, and the ESSA Evidence Level.

Table 1. Benchmark Advance and Benchmark Adelante Studies with ESSA Evidence Level 2 Moderate

Study Number	<i>Benchmark Advance and Benchmark Adelante Study Title</i>	Short Title	ESSA Evidence Level
1	Smarter ELA/Literacy Results for Students in Indian River School District: Progress After Two Years of <i>Benchmark Advance</i> Instruction	Indian River Study 2018	Moderate
2	Smarter ELA/Literacy Results for Students in Indian River School District and the State of Delaware After Initial Year of <i>Benchmark Advance</i> Instruction	Indian River Study	Moderate
3	Comparison of Dual Language Immersion and Monolingual Instruction Using <i>Benchmark Advance</i> and <i>Benchmark Adelante</i>	Chula Vista Study #1	Moderate
10	STAR Early Literacy Growth Achieved After Initial School Year of <i>Benchmark Advance</i> ©2022 Instruction	Grades K-2 Study	Moderate

Table 2. Benchmark Advance and Benchmark Adelante Studies with ESSA Evidence Level 3 Promising

Study Number	<i>Benchmark Advance and Benchmark Adelante Study Title</i>	Short Title	ESSA Evidence Level
4	Evidence-Based Research in the Chula Vista (CA) Elementary School District on the Effectiveness of the <i>Benchmark Advance</i> and <i>Benchmark Adelante</i> Comprehensive Literacy Programs	MSA Study #1	Promising
5	Percentage of Students in Met and Above Category More Than Doubled After Two Years of <i>Benchmark Advance</i> and <i>Benchmark Adelante</i> Instruction	Chula Vista Study #2	Promising
6	Advancing to Proficiency: 2018 State Test Results for Students Using <i>Benchmark Advance</i> and <i>Benchmark Adelante</i>	Four-State Study	Promising
7	California <i>Benchmark Advance</i> and <i>Benchmark Adelante</i> Districts CAASPP Results: After Initial Year of Instruction	71 District Study	Promising
8	Comparison of Dual Language Immersion Using <i>Benchmark Adelante</i> and English-Only Instruction in Kindergarten and 1st Grade Classrooms in North Carolina	North Carolina Study #1	Promising
9	<i>Benchmark Advance</i> Florida: An Efficacy Study	Florida Study	Promising

The number of participants in the ten studies ranges from below 100 participants (Chula Vista Study #1, Chula Vista Study #2, and MSA Study #1) to below 1,000 participants (Indian River Study, Indian River Study 2018, and North Carolina Study #1), to below 6,000 participants (Grades K–2 Study), to greater than 30,000 participants (Four State Study, 71 District Study, and Florida Study).

Subgroup calculations are part of seven of the ten studies. In Indian River Study 2018, the nine subgroups include: Coming from a low-income household, Being a member of the Gap Group, English Learner, Students with Disabilities, African American, Hispanic/Latino, White, Female, and Male. In the MSA Study #1, subgroup calculations are included for Low Socio-Economic Status, English Language Learners, and Students with Disabilities. In the Four-State Study, subgroup comparisons include: English Learners, Economically Disadvantaged, Students with Disabilities, Female, Male, American Indian/Alaska Native, Asian, Black/African American, Hispanic/Latino, Two or More Races, and White. For the 71 District Study, subgroup comparisons include: Economically Disadvantaged; Not Economically Disadvantaged; Ethnicities, including Asian, Black/African American, Filipino, Hispanic or Latino, Two or More Races, and White; and Gender, Female and Male. In the North Carolina Study #1, subgroup comparisons include: Female, Male, Black, Hispanic/Latino, White, identification as an English Learner (ESL), and identification of first language or language spoken at home (English or Spanish). In the Florida Study, subgroup comparisons include: Gender; Economic Status & Minority Status; Students with a Disability; Current English Language Learner; Race/Ethnicity, including American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, Two or More Races, and White. In the Grades K–2 Study, subgroup comparisons include: Gender, SPED Marker, Gifted Marker, ELL Marker, and Race/Ethnicity, including Asian, Black or African American, Hispanic or Latino, American Indian or Alaska Native, Multi Race, and White.

Where statistical significance could be calculated, in five of the studies, it was in favor of the treatment groups. These studies included Indian River Study 2018, Indian River Study, Chula Vista Study #1, North Carolina Study #1, and Grades K–2 Study. Where statistical significance could not be calculated, it was generally due to the use of state test aggregated results where an aggregated standard deviation was not provided or possible to obtain.

Where possible, the effect size was calculated using the Cohen's d formula¹. Effect sizes ranged from medium ($d = .55$ for Grades 3 and 4, and $d = .53$ for Grades 4 and 5 in the Indian River Study) to large ($d = 1.08$ for Grades 3–5 in the Indian River Study 2018) to very large ($d = 1.22$ for English only and $d = 2.03$ for DLI in the Chula Vista Study #1, and for the Grades K–2 Study, Grade K had between $d = 1.534$ and $d = 2.140$), according to Cohen's convention for interpreting effect sizes. In John Hattie's² work on trying to determine what works best in education, he describes an effect size of: 0.5 as equivalent to a one-grade leap; 1.0 as equivalent to a two-grade leap; and above 0.4 as above average for the effects of all educational interventions he has studied. So the effect sizes for these studies, when they could be calculated, are above average.

¹ Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

² <https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/>

Introduction

This report provides a description of Every Student Succeeds Act (ESSA) evidence available, as of December 2023, for Benchmark Education's core elementary literacy and language programs, *Benchmark Advance* and *Benchmark Adelante*. These K–6 programs have been available for general use since the 2016–2017 school year and were used by one district during the 2015–2016 school year while participating in an early study. Bolding was added to emphasize effect sizes, sample sizes, and where there was a control or comparison group(s). The studies summarized in this document are available as complete reports upon request.

Moderate ESSA Evidence

Currently, three of the *Benchmark Advance* and *Benchmark Adelante* studies are quasi-experimental, using existing groups and demonstrating baseline equivalence for those groups. One study is quasi-experimental, using a pretest-posttest design that uses effect sizes and normative score progress³. These four studies qualify for moderate or Level 2 ESSA evidence.

Study 1, Indian River Study 2018

Study title: *Smarter ELA/Literacy Results for Students in Indian River School District: Progress After Two Years of Benchmark Advance Instruction*

- The **treatment group** included students in Grade 3 before *Benchmark Advance* was used as the core ELA program during 2015–2016 school year, and in Grade 5, after two years of instruction with *Benchmark Advance*, during the 2017–2018 school year. There were **854 students in Grade 3** at the pre-test and **876 students in Grade 5** at post-test.
- The control group included students in the same grades who attended schools in the other twelve school districts in Delaware who received ELA instruction with other materials and had no exposure to *Benchmark Advance* materials. There were **6,387 students in the Grade 3** control group at the pre-test, and **6,464 students in the Grade 5** control group at the post-test.
- Using Delaware's Smarter ELA/Literacy Assessment as the outcome measure and using confidence intervals to evaluate the difference between the groups, the treatment and control groups at the pre-test (2016 administration) had confidence intervals that overlapped.
- At post-test, 2018 administration, there was no overlap of the confidence intervals, indicating a **statistically significant difference** between the Indian River students and students in the other twelve school districts in the Grades 3–5 group after the two years of *Benchmark Advance* instruction, **in favor of the treatment group**.
- The size of the difference between the Indian River students and the other twelve school districts was calculated using the Cohen's *d* effect size³. For the all-student treatment group, the effect size was **$d = 1.08$ for the Grades 3–5 students**. For the all-student control group, the effect size was $d = .89$ for the Grades 3–5 students. For the nine subgroups, six of the Indian River subgroups were similar at pre-test with overlapping confidence intervals, and all Indian River subgroups had larger effect sizes than the other twelve-school-district comparison group.

³ Cohen's *d* is defined as the difference between two means divided by the standard deviation of the pooled groups or of the control group alone. For these studies, the pooled standard deviation was used. According to Cohen (1988), an effect size of 0.2 is small, 0.5 is medium, and 0.8 is large. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second edition). New York, NY: Psychology Press, Taylor & Francis Group.

Study 2, Indian River Study

Study title: *Smarter ELA/Literacy Results for Students in Indian River School District and the State of Delaware After Initial Year of Benchmark Advance Instruction*

- The **treatment groups** included students in Grade 3 or 4 during 2015–2016 school year (before new instruction), and in Grade 4 or 5 during the 2016–2017 school year (when the *Benchmark Advance* instruction occurred). There were **854 students in Grades 3–4** at the pre-test, and **853 students in Grades 4–5** at post-test.
- The **control groups** included students in the same grades, Grades 3–4 and Grades 4–5, who attended schools in the other fourteen school districts in Delaware and received ELA instruction with other materials. There were **8,443 students in the Grades 3–4** control group at the pre-test, and **8,358 students in the Grades 4–5** control group at the post-test.
- Using Delaware’s Smarter ELA/Literacy Assessment as the outcome measure and using confidence intervals to evaluate the difference between the groups, the treatment and control groups at the pre-test (2016 administration) had confidence intervals that overlapped.
- At post-test (2017 administration), there was no overlap of the confidence intervals, indicating a **statistically significant difference** between the Indian River students and students in the other fourteen school districts in both the Grades 3–4 and Grades 4–5 groups after the initial year of *Benchmark Advance* instruction, **in favor of the treatment groups**.
- The size of the difference between the Indian River students and the other fourteen school districts was calculated using the Cohen’s *d* effect size. For the treatment groups, effect sizes were **$d = 0.55$ for the Grades 3–4 students** and **$d = 0.53$ for the Grades 4–5 students**. For the control groups, effect sizes were $d = 0.45$ for the Grades 3–4 students and $d = 0.40$ for the Grades 4–5 students.

Study 3, Chula Vista Study #1

Study title: *Comparison of Dual Language Immersion and Monolingual Instruction Using Benchmark Advance and Benchmark Adelante*

- This study was an extension of the original Chula Vista study (MSA Study #1), with additional information from the district to identify students in the different types of instruction and includes two years of instruction using *Benchmark Advance* and/or *Benchmark Adelante*.
- **Three groups** were considered in this study: students receiving **Dual Language Immersion (DLI) instruction (n = 42)**, students receiving **English-only instruction (n = 31)**, and all students in the **Chula Vista Elementary School District**.
- There was no statistically significant difference between the DLI and English-only groups at the pre-test, 2015 administration of the California Assessment of Student Performance and Progress (CAASPP). Aggregated CAASPP results for the Chula Vista Elementary School District made statistical comparison impossible.
- There was a **statistically significant difference** between these groups of students at the post-test, with an **effect size of $d = 1.22$ for the English-only group** and **$d = 2.03$ for the DLI group**.

Study 10, Grades K–2 Study

Study title: *STAR Early Literacy Growth Achieved After Initial School Year of Benchmark Advance ©2022 Instruction*

- This study looked at the first year of implementation of a new ELA core program, *Benchmark Advance* ©2022. The study used a quasi-experimental approach by using a pretest-posttest design with normative score progress from the STAR Early Literacy. The statistical methods included a paired-sample t-test that provided effect sizes also.
- There were **5,199 students** with the STAR Early Literacy scores at BOY, MOY, and EOY. Growth from BOY to EOY for all students, students in each grade level, and all demographic groups had a **statistically significant increase ($p = .001$)**. **Effect sizes were large ($d = 0.747$) to very large ($d = 2.194$)**, based on Cohen's d .
- Based on expected growth from BOY to EOY for the STAR Early Literacy scores, all students, students in each grade level, and all demographic groups **exceeded the expected growth**.
- Looking at the District Benchmark Category Name of **At/Above Benchmark**, at **BOY 26.47%** of students were in this category. At **EOY, 44.22%** of students were in this category.

Promising ESSA Evidence

Currently, there are six *Benchmark Advance* and *Benchmark Adelante* studies that fit into the promising ESSA evidence category. Two studies are related to the Chula Vista Elementary School District school participating in the initial *Benchmark Advance* and *Benchmark Adelante* study, designed and conducted by Main Street Academix (MSA); two studies are related to the many districts in California, Colorado, Michigan, and Minnesota that adopted *Benchmark Advance* and *Benchmark Adelante* since the 2016–2017 school year; and North Carolina Study #1 is related to the comparison of Spanish instruction using *Benchmark Adelante* to English instruction in Kindergarten and Grade 1 classrooms in North Carolina. The newest study to be added to this document is the Florida Study, which is related to the first year of implementation of *Benchmark Advance* ©2022 with Florida standards.

Study 4, MSA Study #1

Study title: *Evidence-Based Research in the Chula Vista (CA) Elementary School District on the Effectiveness of the Benchmark Advance and Benchmark Adelante Comprehensive Literacy Program*

- The participants in this study are from **School A Elementary School in Chula Vista** and are compared to the growth of students in the **other elementary schools in Chula Vista** and **students across the state of California**. During the 2015–2016 school year, there were **98 students in Grade 4 and 93 students in the Grade 5** that received *Benchmark Advance* and/or *Benchmark Adelante* instruction in School A.
- Using the CAASPP as the outcome measure, from pre-test (2015 CAASPP administration) to post-test (2016 CAASPP administration), the Grade 4 students went from 28 percent of students in the Met or Exceeded performance categories on the CAASPP to 51 percent of students in Met or Exceeded, **a 23 percentage point increase** after receiving *Benchmark Advance* and/or *Benchmark Adelante* instruction. The comparison groups had an 11 percentage point increase for students in other Chula Vista elementary schools and a 6 percentage point increase for students across the state of California.

- For the Grade 5 students, School A students went from 31 percent to 70 percent in the Met or Exceeded performance categories, a **39 percentage point increase**. The Grade 5 comparison groups had an 11 percentage point increase for student in other Chula Vista elementary schools and a 9 percentage point increase for students across the state of California.

Study 5, Chula Vista Study #2

Study title: *Percentage of Students in Met and Above Category More Than Doubled After Two Years of Benchmark Advance Instruction*

- This study provides results for the second academic year of instruction with *Benchmark Advance* and/or *Benchmark Adelante*, during the 2016–2017 school year.
- Students included in this report were in Grade 3 prior to *Benchmark Advance* and/or *Benchmark Adelante* instruction in the 2014–2015 school year, and in Grade 5 after two years of *Benchmark Advance* and/or *Benchmark Adelante* instruction in the 2016–2017 school year.
- These **students started at 28 percent (n = 95)** in the Met or Exceeded performance categories on the 2015 CAASPP administration, **grew to 51 percent (n = 98)** in the Met or Exceeded after one year of instruction, and **reached 62 percent (n = 85)** in the Met or Exceeded performance categories by the 2017 CAASPP administration.

Study 6, Four-State Study

Study title: *Advancing to Proficiency: 2018 State Test Results for Students Using Benchmark Advance and Benchmark Adelante*

- By the 2017–2018 school year, **over 180 school districts in four states** used *Benchmark Advance* and/or *Benchmark Adelante* as the core English/Spanish Language Arts curriculum in Kindergarten through either Grade 5 or 6 across all or almost all elementary schools in the districts.
- The districts using the Benchmark Education Company programs, identified as **BEC Districts**, **represent over 500,000 students** receiving ELA/SLA instruction using one or both programs. The districts who were not using Benchmark Education Company programs, identified as Non-BEC Districts, included over 1.6 million student receiving instruction using other materials.
- The state tests in California, Colorado, Michigan, and Minnesota do not report an aggregated standard deviation, so calculating statistical significance and effect sizes are not possible. This study looks at the movement of students into the proficient categories on the state tests from the four states. The growth of the percentage of **students in the proficient categories increased by 3.53 percentage points for BEC Districts** and 1.39 percentage points for Non-BEC Districts from the 2017 to 2018 administrations of the state tests.
- The subgroups included in this study, including the BEC Districts' and Non-BEC Districts' growth in the percentage of student in the proficient categories, include students who are: Identified as **English Learners (BEC Districts up 2.15, Non-BEC Districts up 0.87)**; **Economically Disadvantaged (BEC Districts up 4.19, Non-BEC Districts up 2.34)**; **Students with Disabilities (BEC Districts up 2.51, Non-BEC Districts up 1.63)**; **Female (BEC Districts up 3.47, Non-BEC Districts up 1.20)**; **Male (BEC Districts up 3.58, Non-BEC Districts up 1.68)**; **American Indian/Alaska Native (BEC Districts up 3.58, Non-BEC Districts up 1.45)**; **Asian (BEC Districts up 2.11, Non-BEC Districts up 1.39)**; **Black/African American (BEC Districts up 3.13, Non-BEC Districts up 0.54)**; **Hispanic/Latino (BEC Districts up 3.76, Non-BEC Districts up 3.07)**; **Two or More Races (BEC Districts up 1.85, Non-BEC Districts up 1.69)**; and **White (BEC Districts up 1.89, Non-BEC Districts up 0.33)**.

Study 7, 71 District Study

Study title: *California Benchmark Advance Districts CAASPP Results: After Initial Year of Instruction*

- During the initial year *Benchmark Advance* and *Benchmark Adelante* were available for adoption in California, 71 districts committed to a long-term relationship with Benchmark Education. The students in the 71 districts represent about 6.5 percent of students enrolled in California elementary schools.
- The Grades 3–4 group in the 71 districts contained **30,184 students tested on the 2016 CAASPP** and **29,950 students tested on the 2017 CAASPP**. The Grades 3–4 comparison group in the other California districts had **402,855 students tested on the 2016 CAASPP** and **401,296 students tested on the 2017 CAASPP**.
- The Grades 4–5 group in the 71 districts contained **31,069 students tested on the 2016 CAASPP** and **30,871 students tested on the 2017 CAASPP**. The Grades 4–5 comparison group in the other California districts had **417,999 students tested on the 2016 CAASPP** and **415,791 students tested on the 2017 CAASPP**.
- The CAASPP does not report an aggregated standard deviation, so calculating statistical significance or effect sizes is not possible. But this study looks at the movement of students into the Met and Above category on the CAASPP. **For Grades 3–4**, the 71-districts group started at 42.77 percent and move to 46.44 percent, **an increase of 3.67 percentage points**; the other California districts started at 42.45 percent and moved to 45.03, an increase of 2.58 percentage points. **For the Grades 4–5**, the 71-districts group started at 44.52 percent and move to 48.24 percent, **an increase of 3.72 percentage points**; the other California districts started at 43.74 percent and moved to 46.48, an increase of 2.74 percentage points. There is about one percentage point difference in growth for both grade groups, favoring students in the 71 districts.
- This California report also includes the analysis of the following **demographic subgroups: Economically Disadvantaged; Not Economically Disadvantaged; Ethnicities, including Asian, Black or African American, Filipino, Hispanic or Latino, Two or More Races, and White; and Gender, Female and Male.**

Study 8, North Carolina Study #1

Study title: *Comparison of Dual Language Immersion Instruction Using Benchmark Adelante and English-Only Instruction in Kindergarten and Grade 1 Classrooms in North Carolina*

- The focus of this study is to compare the results of instruction using *Benchmark Adelante* in a Dual Language Immersion (DLI) classroom with English-Only (EO) classrooms using English materials (but not *Benchmark Advance*) in Kindergarten and Grade 1 classrooms in three schools in a district located in the eastern part of North Carolina.
- A total of 127 students, **64 in Kindergarten and 63 in Grade 1, received instruction using Benchmark Adelante** and assessments in Spanish (IDEL). There were **590 students (283 in Kindergarten and 307 in Grade 1) who received English instruction** and assessments in English (DIBELS). IDEL and DIBELS measures that assessed the same basic early literacy skills were compared.

- Looking at the effect sizes from initial to final scores, measuring **phonemic awareness in Kindergarten (using FSF and PSF), for DLI students $d = 2.44$** and for EO students $d = 1.59$; measuring the **alphabetic principal in Kindergarten (using FPS and NWF), for DLI students $d = 1.42$** and for EO students $d = 0.95$; measuring the **alphabetic principal in Grade 1 (using FPS and NWF), for DLI students $d = 1.66$** and for EO students $d = 1.30$; and measuring **accurate and fluent reading of connect text in Grade 1 (using FLO and DORF), for DLI students $d = 0.89$** and for EO students $d = 0.75$. Scores are **statistically significantly different from the initial to final** scores across the measures above.
- This report also includes the analysis of the demographic subgroups including, among other categories, **English Learners** or not and **home language of English or Spanish**.

Study 9, Florida Study

Study title: *Benchmark Advance Florida: An Efficacy Study*

- The initial year *Benchmark Advance* was available for adoption in Florida was during the 2021–2022 school year. Of the 67 districts in Florida, **24 were using the *Benchmark Advance* Florida version (BA FL)** or would be fully using the program during the next school year. The BA FL group was compared to the **43 school districts (Non-BEC)** using other English Language Arts materials during the same period.
- The Florida Standards Assessment (FSA) results from the spring administration of 2021 and 2022 were used in this study.
- For BA FL, two cohorts were examined in this study: **Grade 3 to 4 ($n = 77,195$ for the spring 2021 FSA administration and $n = 77,200$ for the spring 2022 FSA administration)** and **Grade 4 to 5 ($n = 77,623$ for the spring 2021 FSA administration and $n = 81,578$ for the spring 2022 FSA administration)**. For Non-BEC districts two cohorts were examined in this study Grade 3 to 4 ($n = 121,173$ for the spring 2021 FSA administration and $n = 120,234$ for the spring 2022 FSA administration) and Grade 4 to 5 ($n = 124,068$ for the spring 2021 FSA administration and $n = 129,665$ for the spring 2022 FSA administration).
- This study presents the analysis of change in the percent of students who reached the Level 3 & Above performance categories on the FSA from the spring 2021 to spring 2022. For all students in the Grade 3 to 4 groups, the **BA FL group gained 3.22**, the **Non-BEC group gained 2.22**, and the **state of Florida gained 2.59**. For all students in the Grade 4 to 5 groups, the **BA FL group gained 3.24**, the **Non-BEC group gained 2.18**, and the **state of Florida gained 2.59**.
- This study also includes the analysis of the following demographic subgroups: Gender; Economic Status & Minority Status; Students with a Disability (SWD); Current English Language Learners (ELL); Race/Ethnicity, including American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, Two or More Races, and White.

Summary

Benchmark Education will continue to look for research opportunities, including but not limited to randomized controlled trials. Instead of putting all eggs in one basket, we will continue to look at almost any research opportunity as a good one, including working with: districts who want to understand how students are benefiting from *Benchmark Advance* instruction; classroom teachers who are in graduate school and interested in doing action research; state test results; and organizations and universities who are applying for federal grants and need partners. While the evidence structure provided by the “Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments”⁴ clarifies the definitions provided in section 8101(21)(A) of ESSA, it does not suggest the restriction to only strong evidence.

⁴ <https://www2.ed.gov/policy/elsec/leg/essa/guidanceusesinvestment.pdf>

