

RESEARCH RESULTS

HMH Into Literature®: A Quasi-Experimental Study

STUDY PROFILE

SAMPLE:

807 schools in Texas (252 Treatment; 555 Control)

GRADES:

6–8

EVALUATION PERIOD:

Spring 2021; Spring 2022

OUTCOME MEASURES:

State of Texas Assessment of Academic Readiness (STAAR®)

STUDY CONDUCTED BY:

JEM & R LLC

STUDY DESIGN:

Quasi-experimental design

ESSA EVIDENCE LEVEL:

Moderate

THE CHALLENGE

As educators recover from the disruptions caused by the COVID-19 pandemic and work to accelerate literacy instruction for middle-school students, it is essential to provide effective English language arts instruction that focuses on fluency as well as reading comprehension. Furthermore, given the significant and widespread unfinished learning that resulted from the COVID-19 pandemic, it is essential that educational programs provide research-based instruction and be examined to determine the extent to which they aid learning recovery efforts.

THE SOLUTION

Based on evidence-based principles, *HMH Into Literature* is a comprehensive English language arts program that offers rich content, actionable insights, personalized learning, and standards-based instruction—all within one seamless experience. High-interest, relevant materials engage students, and Reader's Choice and self-serve resources support student agency, choice, and growth mindset. For teachers, *Into Literature* provides a flexible design, including expanded access to rich and varied digital resources for each literacy strand and differentiation tools for multilingual, striving, and advanced learners. The program provides the instructional tools, rich pedagogy, and professional services to ensure that teachers and students not only reach but also exceed their instructional goals.

THE STUDY

Houghton Mifflin Harcourt (HMH) contracted with an independent research and evaluation firm, JEM & R, to determine the effectiveness of its *Into Literature* program in helping students gain literacy skills. The quasi-experimental study uses recent Texas state assessment data from Spring 2021 and 2022 to determine the relationship between *Into Literature* and grade 6–8 student English language arts (ELA) performance. The study included 252 schools that purchased and used the latest edition of *Into Literature* and 555 closely matched comparison schools.

SCHOOL CHARACTERISTICS

The study's sample included Texas middle schools with students in grades 6–8. The study included 252 schools that purchased and began using *Into Literature* during the 2019–2020 school year (Cohort A) or 2020–2021 school year (Cohort B). Confirmation calls were conducted with schools/districts on the HMH sales list to determine the extent to which identified schools have used *Into Literature*. These confirmation calls allowed researchers to determine that potential treatment schools were established *Into Literature* users. Demographic characteristics of schools are shown in Table 3.

IMPLEMENTATION OVERVIEW

Given the nature of the study, researchers relied on self-report measures from confirmation calls to determine usage of *Into Literature* in English language Arts classes in grades 6–8. The following data was collected: (1) verification of use of the *Into Literature* program and at which grades, and (2) the proportion of students within schools who used this curriculum.

Across all subgroups analyzed in both Post Year 2 and Post Year 3 including Whites, Hispanics, African Americans, students with disabilities, economically disadvantaged students, and English learners, Into Literature students demonstrated higher proficiency rates on average as compared to non-users.

PARTICIPANTS

The study consisted of 252 *Into Literature* schools and 555 closely matched schools that served as the comparison group. Because performance data is available for each grade level, the unit of analysis is *grade-level performance* within schools. Therefore, *Into Literature* schools were matched by school-level characteristics and grade-level performance measures to unique schools within the state.

Two cohorts are included in the analytical samples as outlined in Table 1. Cohort A consists of Texas middle schools that began using *Into Literature* in the 2019–20 school year, while Cohort B consists of Texas middle schools that began using *Into Literature* in the 2020–21 school year. Of note, state assessment data is not available for the 2019–20 school year. As such, 2018–19 serves as the baseline year for both cohorts, although Cohort B's true baseline corresponds with the 2019–20 school year (during the pandemic when state tests were waived).

TABLE 1: INTO LITERATURE'S IMPACT ANALYSIS COHORTS

Academic Year	Cohort A	Cohort B
2018–19	Baseline	Baseline
2019–20	Post Year 1*	--*
2020–21	Post Year 2**	Post Year 1
2021–22	Post Year 3	Post Year 2**

*No state assessment data available

**These schools are included in the Full Sample analyses.

Table 2 displays the final grade-level units for treatment and comparison schools included in the final analytic sample.

TABLE 2: ANALYTIC SAMPLE SIZE BY GRADE AND COHORT

Cohort	State	Grade Level			Total
	Group	6	7	8	
A (2019–20)	Total	410	394	394	1198
	Control	205	197	197	599
	<i>Into Literature</i>	205	197	197	599
B (2020–21)	Total	20	44	44	108
	Control	10	22	22	54
	<i>Into Literature</i>	10	22	22	54

MEASURES

STATE OF TEXAS ASSESSMENTS OF ACADEMIC READINESS (STAAR) PROGRAM

The State of Texas Assessments of Academic Readiness (STAAR) program has been in use since Spring 2012 to measure the Texas Essential Knowledge and Skills (TEKS) curriculum standards in math, reading, and language arts in grades 3–8, as well as an end-of-course assessment for Algebra I, English I and English II. The tests are vertically scaled in grades 3–8 to allow for direct comparison of student test scores across grade levels within a content area.

The present study uses ELA assessment data from the STAAR. STAAR performance standards relate levels of test performance to the expectations defined in the TEKS. Cut scores established by the agency distinguish between performance levels, or categories (Masters Grade Level, Meets Grade Level, Approaches Grade Level, and Did Not Meet Grade Level). For the present study, percent proficient (Masters Grade Level and Meets Grade Level) is the student achievement measure.

PROPENSITY MATCH PROCEDURES

To evaluate the impact of *Into Literature*, a comparison group was identified using Propensity Score Matching (PSM). PSM was conducted separately for each grade level within each school, and within each cohort. To create these comparison groups, researchers conducted a three-step process:

STEP 1: COMPUTING PROPENSITY SCORES

For each unit included in a model, the probability of receiving the *Into Literature* treatment was estimated using a logistic regression model, with the *Into Literature* treatment flag as the outcome variable and 11 grade-level characteristics (e.g., number of students, percent English learners) as predictor variables.

STEP 2: MATCHING

PSM was implemented with MatchIt version 4.5.0 (Ho et al., 2011) using the nearest neighbor-matching method, with generalized linear model specified as the distance measure and logit specified as the link function. In each PSM model, each *Into Literature* unit was sequentially matched with one potential control unit.

STEP 3: ASSESSING BALANCE

Baseline equivalence was assessed on key characteristics using a two-sided Welch's Two Sample t-test with a 0.05 alpha-level.

Demographic characteristics of the matched schools are shown in Table 3.

TABLE 3: DESCRIPTIVE STATISTICS FOR THE FULL SAMPLE (MEANS AND STANDARD DEVIATIONS)			
Group	Category	<i>Into Literature</i>	Control
Avg. Baseline Proficiency	--	43.67% (17.61)	42.98% (16.53)
Avg. # of Students by Grade	6	215	215
	7	219	219
	8	219	219
Gender (% in School)	Male	50.10% (11.43)	50.25% (9.30)
	Female	47.15% (11.07)	47.61% (8.99)
Race/Ethnicity (% in School)	White	28.03% (25.12)	27.35% (26.85)
	African American	14.52% (16.47)	15.24% (19.67)
	Hispanic	52.45% (24.45)	52.50% (29.31)
	Asian	2.53% (4.49)	2.30% (4.54)
	Two or more races	2.00% (1.68)	2.11% (1.89)
Subpopulations (% in School)	Economically Disadvantaged	64.77% (23.98)	65.05% (24.95)
	English Learners	20.70% (18.33)	20.77% (19.95)
	Students With Disabilities	8.81% (4.23)	8.60% (5.03)

POST YEAR 3 ANALYSES

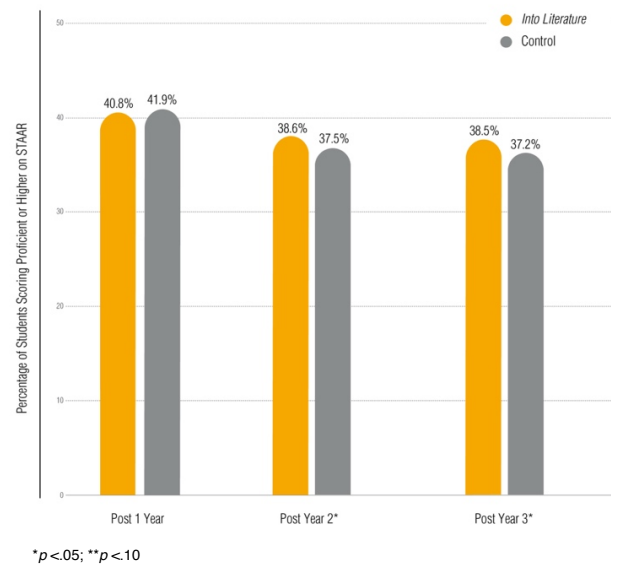
Analysis of the main effects of Cohort A Post Year 3 (i.e., Spring 2022) also show that a statistically significant difference was observed between *Into Literature* schools and non-*Into Literature* schools in their reading performance following three years of program usage, $p < .05$. As shown in Figure 1, *Into Literature* schools had a higher proficiency rate by 1.2% points than schools that did not use *Into Literature*.

POST YEAR 1 ANALYSES

Post Year 1 data was only available from Cohort B; this consisted of 54 schools in Texas that began using *Into Literature* in 2020–21 (new users of the program). Thus, Post Year 1 represents reading proficiency rates from Spring 2021, approximately 8 months following initial implementation.

Results from this analysis indicate that while the Post Year 1 proficiency rate for matched comparison schools was 1.2% points higher than *Into Literature* users, these differences were not statistically significant, $p > .05$, see Figure 1.

FIGURE 1. ESTIMATED MARGINAL MEANS OF INTO LITERATURE AND CONTROL SCHOOLS – OVERALL POST-YEAR MAIN EFFECTS



RESULTS

HOW DOES STUDENT ACHIEVEMENT IN LITERACY DIFFER ACROSS USERS AND NON-USERS OF INTO LITERATURE?

To examine the effectiveness of *Into Literature*, analyses compared the performance of schools using *Into Literature* relative to schools that did not use *Into Literature* for each post year separately. All analyses (ANCOVA) controlled for 2019 math proficiency rates to ensure equivalency across groups and increase the sensitivity of analyses.

To maximize the sample size available for analyses of main effects, schools in Cohort A and B were combined to examine the impact of the program in Post Year 2, the only common post-year data available from both cohorts. Post Year 1 program effects were estimated from Cohort B only (with Post Year 1 data), and Post Year 3 program effects were estimated from Cohort A only (with Post Year 3 data).

FULL SAMPLE (POST YEAR 2) ANALYSES

Post Year 2 consists of Spring 2021 data from Cohort A (the largest sample) and Spring 2022 data from Cohort B. Results from the full, combined sample are presented in Figure 1.

Overall results of main effects show that schools using the *Into Literature* curriculum for two years demonstrated a significantly higher proficiency rate than non-users, $p < .05$. The percent of students proficient in reading was 1.1% points higher among *Into Literature* students in Texas as compared to control schools.

ANALYSES BY YEAR

To look at historical trends irrespective of years of usage, data from both cohorts for each year following COVID-related school closures were examined. Spring 2022 data consisted of data from Cohort A Post Year 3 and Cohort B Post Year 2, and Spring 2021 data consisted of proficiency rates from Cohort A Post Year 2 and Cohort B Post Year 1.

Results (presented in Figure 2 below) demonstrated marginally significant differences in Spring 2022 and Spring 2021 proficiency rates in favor of *Into Literature*, $p < .10$. In both cases, proficiency rates for schools that used *Into Literature* were 1% point higher than for matched comparison schools.

GRADE-LEVEL ANALYSES

To examine possible grade-level differences between students attending *Into Literature* and control schools, exploratory analyses examined Post Year 2 and 3 performance among students in grades 6–8. Given the small sample sizes for Cohort B (Post Year 1), analyses were not run for this sub-sample.

Results for Post Year 2 (full sample) are presented in Figure 3. Analysis by grade levels showed positive patterns in favor of *Into Literature*. At grade 6, there was a significant difference following two years of program usage. Proficiency rates for schools using *Into Literature* was 1.9% points higher than for non-users, $p < .05$. At grades 7 and 8, schools using *Into Literature* also demonstrated higher proficiency rates than did non-users; however, these differences were not significant, $p > .05$.

Results were consistent at Post Year 3 (Cohort A). After three years of program usage, grade 6 students demonstrated a marginally significant difference, with *Into Literature* students performing 1.8% points higher than students using other literacy programs, $p < .10$. No significant differences were observed for grades 7 and 8 at Post Year 3.

FIGURE 2. ESTIMATED MARGINAL MEANS OF INTO LITERATURE AND CONTROL SCHOOLS – SPRING 2022 MAIN EFFECTS

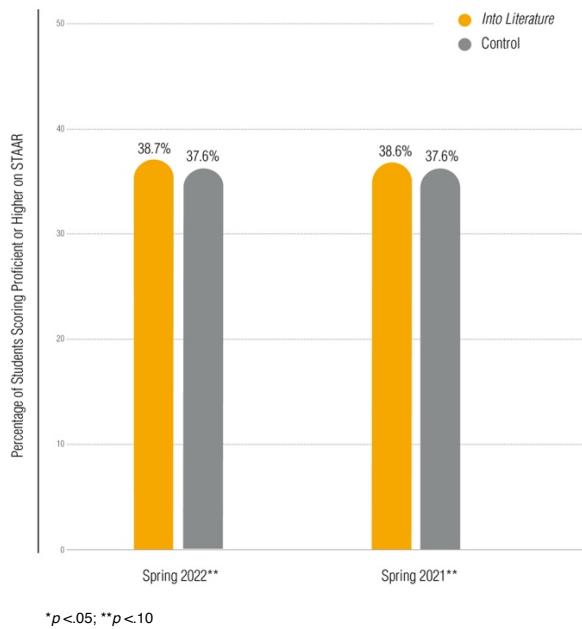
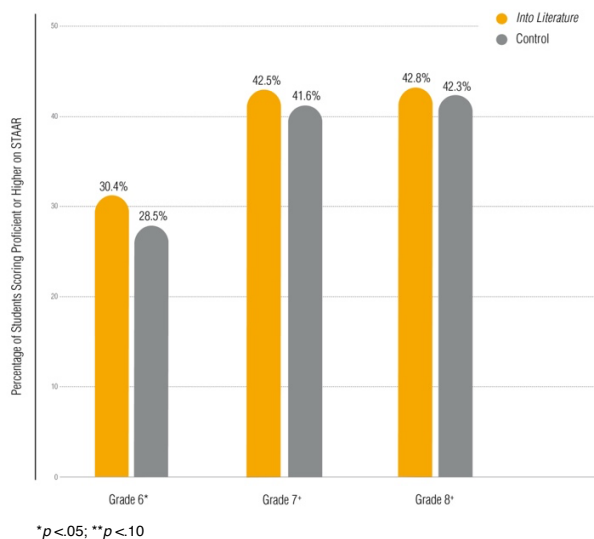


FIGURE 3. ESTIMATED MARGINAL MEANS OF INTO LITERATURE AND CONTROL SCHOOLS – POST YEAR 2 RESULTS BY GRADE



IS INTO LITERATURE ASSOCIATED WITH SIGNIFICANT IMPACTS FOR VARIOUS SUBPOPULATIONS OF STUDENTS?

To examine subgroup-specific differences between students attending *Into Literature* and control schools, exploratory¹ analyses examined Post Year 2 and 3 performance among students in different subpopulations. Separate analyses were conducted by ethnicity (White, Hispanic, African American), students with disabilities, English learners, and economically disadvantaged students².

To maximize the sample size, all comparisons were made using the overall sample (collapsed across grades). It should be noted that for some comparisons, sample sizes are quite small; therefore, results should be interpreted with caution.

ETHNICITY

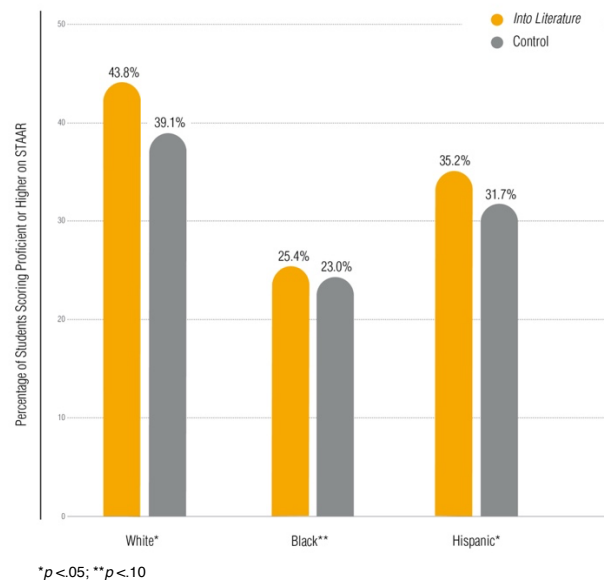
Comparisons among different racial and ethnic groups showed statistically significant differences for White (Post Year 2 and Year 3), Black (Post Year 3), and Hispanic (Post Year 2 and Year 3) students, as shown in Figure 4.

White students using *Into Literature* outperformed those in control schools, and this was statistically significant at Post Year 2 and Year 3, $p < .05$. Specifically, ELA performance among White students using *Into Literature* was 4.7% and 5.7% points higher than non-users in Post Year 2 and Year 3, respectively.

Among Hispanic students, *Into Literature* students outperformed control students by approximately 3.5% points and 4.2% points respectively, $p < .05$.

With respect to African American students, *Into Literature* users demonstrated higher proficiency rates across both Year 2 and Year 3; a marginally significant difference was observed for Post Year 2, and a significant difference was seen for Post Year 3.

FIGURE 4. ESTIMATED MARGINAL MEANS OF INTO LITERATURE AND CONTROL SCHOOLS – POST YEAR 2 RESULTS BY ETHNICITY



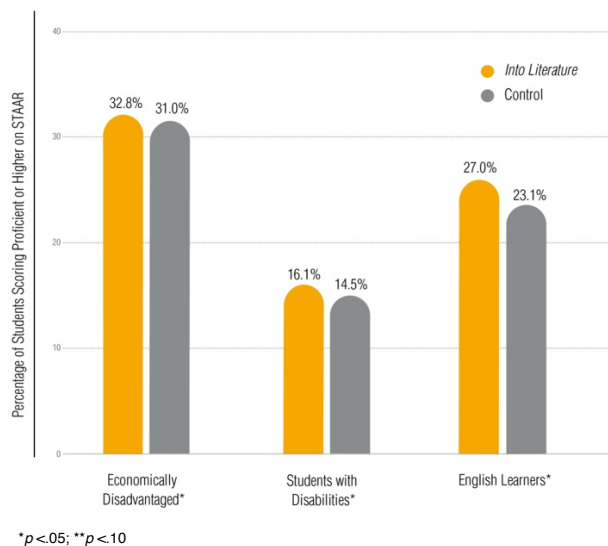
¹ The subgroup effects are viewed as exploratory and need theoretical frameworks and other rigorous experimental designs in the future to be estimated "causally." Further, analyses are based on smaller sample sizes.

² Analyses by gender were not possible because this data was not available from the Texas Education Agency.

OTHER SUBGROUPS

Subgroup comparisons among students with disabilities, economically disadvantaged students, and English learners at Post Year 2 showed that *Into Literature* users demonstrated higher ELA proficiency rates. Specifically, significant differences were observed across Post Year 2 for students classified as economically disadvantaged, students with disabilities, and English learners. These differences can be observed in Figure 5. For Post Year 3, significant differences were observed among economically disadvantaged students and English learners, $p < .05$.

FIGURE 5. ESTIMATED MARGINAL MEANS OF INTO LITERATURE AND CONTROL SCHOOLS – POST YEAR 2 RESULTS BY SUBGROUP



CONCLUSION

Results from the *Into Literature* quasi-experimental study showed that *Into Literature* usage is associated with positive effects in student ELA performance in Texas.

Into Literature schools demonstrated higher proficiency rates than control schools following two and three years of program use; statistically significant differences were observed in Post Year 2 and Post Year 3 ($p < .05$). Additionally, marginally significant differences were observed for Spring 2022 and Spring 2021 ($p < .10$), irrespective of the number of years of program

usage.

Examination of results within each grade level also showed a similar positive pattern of results. In general, *Into Literature* users demonstrated higher proficiency rates within all grade levels, and results were statistically significant for grade 6 in Year 2 and marginally significant for grade 6 in Year 3. Across all subgroups analyzed in both Post Year 2 and Post Year 3 including Whites, Hispanics, African Americans, students with disabilities, economically disadvantaged students, and English learners, *Into Literature* students demonstrated higher proficiency rates on average as compared to non-users. Furthermore, in Year 2, significant differences were observed for all subgroups except one (African American students were marginally significant).

Such consistency in positive trends across multiple grades and subpopulations increases confidence that *Into Literature* has a positive impact on student ELA performance.

That said, the study has several limitations that should be considered when interpreting results. These include the following: a) as a quasi-experiment, it is possible that other factors may be associated with differences that cannot be ruled out; b) given masking of data due to small sample sizes within subgroups, data do not represent all impacted students; c) for Cohort B, spring 2022 is the first "post" year of data available, and these schools had used *Into Literature* for only approximately 7–8 months (state testing occurs in April); and d) there is no implementation data on how well teachers implemented the program and their level of fidelity to the learning model. All of these factors can diminish the detection of effects.

Despite these limitations, however, the results from this quasi-experimental study using Texas state assessment data provide evidence that *Into Literature* is an effective ELA program that can help accelerate student learning.

REFERENCES

Ho, D., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1–28. <https://doi.org/10.18637/jss.v042.i08>.

Check out more *Into Literature* research at hnhco.com/researchlibrary



Houghton Mifflin Harcourt.

hnhco.com

HMH Into Literature®, Houghton Mifflin Harcourt® and HMH® are registered trademarks of Houghton Mifflin Harcourt. © Houghton Mifflin Harcourt. All rights reserved. 06/23