THE EDTECH COLLECTIVE
Instructure Partner Ecosystem

WAYGROUND

ESSA Level III Brief (2023–24)
Implementation and Effectiveness in Texas

Andrew Scanlan, M.A., Senior Researcher
Elizabeth Allen Green, Ph.D., Researcher

May 27, 2025

Main Research Findings	
Reading	
+	School-level usage of *Wayground*[1] (the total number of student responses across all grades) showed a **significant positive relationship with 3rd—8th grade reading outcomes**
Mathematics	
+	School-level usage of *Wayground* (the total number of student responses across all grades) showed a **significant positive relationship with 3rd—8th grade math outcomes**

INTRODUCTION

Wayground provides an instructional suite where educators can create and deliver accessible curriculum resources intended to meet every student's needs across all grade levels and subjects.

Wayground contracted with Instructure, a third-party edtech research company, to examine the relationship between the total number of student responses completed on its platform at schools and average reading and math learning outcomes at each school. Using the Every Student Succeeds Act (ESSA) standards as guidance in developing a study design, findings in this report align with ESSA Level III (Promising Evidence) (see Appendix A).

[1] *Wayground* was formerly known as *Quizizz*.

RESEARCH QUESTIONS

Implementation

1. On average, how many responses did students complete on *Wayground* at each school?[2]

Student outcomes

2. On average, did 3rd–8th grade students perform better on the State of Texas Assessments of Academic Readiness (STARR) reading and math assessments in schools with more *Wayground* usage?

STUDY DESIGN AND METHODS

This study used a correlational design—aligned with ESSA Level III evidence standards—to examine publicly available school-level data provided by the Texas Education Agency. It included 6,646 public schools. Spring 2024 STARR reading and math assessments, which are completed by all Texas public school students in 3rd–8th grade annually, served as the student outcomes for the study. To mitigate bias, the study included the following school-level controls: weighted average school-level spring 2023 STARR reading and math scale scores;[3] school district; total enrollment; as well as the percentages of students by race/ethnicity categories, special education status (SPED), and Title I status (see more in Appendix B).

Researchers conducted a multilevel modeling (MLM) analysis to examine the relationship between school-level *Wayground* usage and student performance on the spring 2024 STAAR reading and math assessments.[4] *Wayground* usage was categorized using *k*-means clustering to group schools into four levels: non-use, low, medium, and high. The "non-use" group (n = 639) included schools that had no student responses submitted on *Wayground* during the 2023–24 school year. While this group served as the reference category in the analysis, it is important to note that baseline equivalence between groups was not assessed. As such, findings should be interpreted as correlational rather than causal.

IMPLEMENTATION FINDINGS

During the 2023–24 school year, students in these schools completed 3,485,477 total responses on *Wayground* across all grades.[5] On average, students in high usage schools (n = 119) submitted 1,143,604 total responses; students in medium usage schools (n = 669) submitted 383,937 total responses and students in low usage schools (n = 5,219) submitted 44,342 total responses (see Table 1).

---

[2] *Wayground* allows for various response formats including text answers, multiple choice, open-ended questions, audio responses, and video responses.
[3] Because STAAR scale scores were only available at the grade level (3rd–8th grade), researchers constructed a weighted average scale score for each school. This was calculated using the number of tests administered per grade and the corresponding average scale scores, serving as a proxy for each school's overall performance.
[4] *Wayground* usage was not available at the grade level.
[5] Schools may have been using *Wayground* for one or multiple school years.

Table 1. *Wayground* average usage by school and usage group (non-use group not shown)

Usage Group	Average Number of Student Responses	Standard Deviation	Range of Responses
High usage (n = 119)	1,143,604	439,395	767,512 to 3,485,477
Medium usage (n = 669)	383,937	137,287	214,838 to 761,205
Low usage (n = 5,219)	44,342	49,523	2 to 213,552

STUDENT OUTCOMES FINDINGS

For **reading**, the total number of student responses on *Wayground* at each school (across all grades) showed a significant positive relationship with weighted average 3rd–8th grade scale scores at the low (z = 0.07 p < .001),[6] medium (z = 0.15 p < .001), and high (z = 0.17 p < .001) usage levels.

For **math**, the total number of student responses on *Wayground* at each school (across all grades) showed a significant positive relationship with weighted average 3rd–8th grade scale scores at the low (z = 0.07 p < .001), medium (z = 0.15 p < .001), and high (z = 0.18 p < .001) usage levels (see Appendix C).

LIMITATIONS AND FUTURE RESEARCH

The current study offers promising results for *Wayground*, but further research is needed to address its limitations and strengthen findings:

- Limited context around the difference between usage and non-usage schools: Future research should use an ESSA Level II quasi-experimental design that establishes baseline equivalence between groups. This study's findings are limited by potential unknown differences between schools with *Wayground* use and those without. There may be meaningful differences in the demographics and characteristics of both groups that affect reading and math performance other than the intervention. As a result, this study's findings are correlational, only, and cannot be said to be representative of the causal impact of *Wayground* usage on outcomes.
- School-level usage: The study analyzed *Wayground* use at the school level using weighted average scale scores. Important relationships between grade-level use and grade-level outcomes may be masked as a result because different grades have distinct curriculum, contexts, and learning needs.
- No student-level data: The study did not analyze individual student-level usage, which limits insights into how specific students (and student subgroups) may engage with and benefit from *Wayground*.
- Limited to Texas: The study was limited to Texas using state assessment data, limiting generalizability. Future research should replicate analyses in other states.

[6] A z-score shows how much higher or lower a value is compared to the average. A z-score of 1 means the value is one standard deviation above average.

CONCLUSIONS

Given the positive findings, this study provides results to satisfy ESSA evidence requirements for Level III (Promising Evidence).

APPENDIX A

The Every Student Succeeds Act (ESSA) provides schools and districts with a framework for determining which products are evidence-based and have been shown to improve student or other relevant outcomes. Following guidance from ESSA (statute and non-regulatory guidance), Education Department General Administrative Regulations (EDGAR), Standards for Excellence in Education Research (SEER) and What Works Clearinghouse (WWC), Instructure classifies the research of interventions into one of the four ESSA evidence levels. For more information regarding the evidence levels, please visit https://www.instructure.com/resources/product-overviews/ensure-edtech-efficacy-essa-evidence.

 **ESSA Level IV** **Demonstrates Rationale**	 **ESSA Level III** **Promising Evidence**	 **ESSA Level II** **Moderate Evidence**	 **ESSA Level I** **Strong Evidence**
Research-based logic model (theory of change) for why this product should work **Blueprint for implementation** with fidelity, including appropriate usage metrics to track Represents a rationale – **not empirical research** – in an authentic education setting **Limitations** on federal funding eligibility	**Correlational research study** showing positive relationship between tool use and student outcomes Study **did not include comparison groups**, random assignment, or baseline equivalence Most meaningful for districts with **similar context** (student demographics, etc.) Establishes **eligibility for all types** of federal funding	**Quasi-experimental research study** showing students who used the product outperformed students who did not Includes **demographically similar comparison group**, but groups were not randomly assigned District **context should be strongly considered** when interpreting results Establishes **eligibility for all types** of federal funding	**Experimental research study** proving students who used the product outperformed students who did not Utilizes **randomized comparison group** for very strong, highly generalizable evidence Establishes **eligibility for all types** of federal funding

APPENDIX B

Table B1: Average student demographics and characteristics at each school in the analytic sample

Demographic Category	Group	Percentage of overall sample
Gender	Female	49.0%
	Male	51.0%
Race/ethnicity	Black	12.3%
	Hispanic	53.3%
	White	26.6%
	American Indian	0.3%
	Pacific Islander	0.1%
	Two or more races	3.2%
Title I		78.6%
Special education		15.5%
English language learners		25.2%
Economically disadvantaged		66.2%
At-risk		53.5%
Enrollment size (count)		564

Note: Demographic categories are rounded so the sum of subcategories may not equal 100%.

APPENDIX C

Researchers conducted a multilevel modeling analysis to examine the relationship between school-level *Wayground* usage and student performance on the spring 2024 STAAR reading and math assessments. *Wayground* usage was categorized using *k*-means clustering to group schools into four levels: non-use, low, medium, and high.

The "non-use" group (*n* = 639) included schools that had no student responses on *Wayground* during the 2023–24 school year. While this group served as the reference category in the analysis, it is important to note that baseline equivalence between groups was not assessed. As such, the findings should be interpreted as correlational rather than causal. Baseline equivalence between user and non-user schools was not examined in this study. There may be meaningful differences in the demographics and characteristics of both groups that affect reading and math performance other than the intervention.

Because STAAR scale scores were only available at the grade level (3rd–8th grade), researchers constructed a weighted average scale score for each school. This was calculated using the number of tests administered per grade and the corresponding average scale scores, serving as a proxy for each school's overall performance. The model accounted for the nested structure of the data, recognizing that schools are situated within districts that may differ in policies and contextual factors influencing student achievement. The analysis also controlled for several statistically significant covariates depending on subject. In reading, these included weighted average spring 2023 STAAR scale score, total school enrollment, and the percentages of White, Black, Hispanic, SPED, and Title I students at each school. In math, these included weighted average spring 2023 STAAR math scale score, total school enrollment, and the percentages of Black, SPED, and Title I students at each school.

Effect size values reflect standardized coefficients (i.e., change in *z*-scored STAAR reading and math outcomes) from a parallel model using the same predictors. These indicate the relative magnitude of difference in standard deviation units, compared to the non-use group. Results meeting the threshold for statistical significance ($p < .05$) are highlighted in **green**.

Table C1: MLM results for reading by usage group using weighted average scale scores

School-level usage group (across all grades)	Unstandardized beta coefficient (raw STARR scores)	Standard Error	*z*-value	*p*-value	Effect Size
Low use (*n* = 5,219)	7.11	1.58	4.49	< .001	**0.07**
Medium use (*n* = 669)	14.47	2.21	6.56	< .001	**0.15**
High use (*n* = 119)	16.22	3.69	4.40	< .001	**0.17**

Table C2: MLM results for math by usage group using weighted average scale scores

School-level usage group (across all grades)	Unstandardized beta coefficient (raw STARR scores)	Standard Error	*z*-value	*p*-value	Effect Size
Low use (*n* = 5,219)	8.87	1.77	5.02	< .001	**0.07**
Medium use (*n* = 669)	18.39	2.48	7.40	< .001	**0.15**
High use (*n* = 119)	21.95	4.14	5.31	< .001	**0.18**