

MISSION.IO

Mission.io and Fifth Grade Learning Outcomes: An ESSA Level II Efficacy Study, 2025–26

Prepared for:
Mission.io

Prepared by Instructure:
Justine Bautista, Ph.D., Instructure Contractor
Molly Henschel, Ph.D., Interim Director of Research
Elizabeth Allen Green, Ph.D., Researcher

June 2026



EXECUTIVE SUMMARY

Mission.io contracted with Instructure to examine the impact of *Mission.io* on students' behavioral engagement, content knowledge, writing performance, disaffection, and writing self-efficacy. Using the Every Student Succeeds Act (ESSA) standards as guidance in the study design, the engagement findings in this report aligns with Level II requirements (Moderate Evidence).

Study Sample and Methodology

This ESSA Level II quasi-experimental study evaluated *Mission.io* implementation among fifth grade students during the 2025–26 school year. The analytic sample included 1,039 fifth grade students across 53 classrooms. The treatment group totaled 582 students who used *Mission.io* and the comparison group totaled 457 students who did not use the platform.

Researchers analyzed *Mission.io*'s implementation, demographic data, usability survey responses, and pretest and posttest outcome measures to assess the platform's impact on students' outcomes. All analyses used outcome-specific, propensity-score weighting with the matched pretest retained as a covariate, and cluster-robust standard errors at the classroom level. Hedges' *g* effect sizes used a small-sample correction. All analyses met What Works Clearinghouse (WWC) Version 5.0 baseline equivalence standards (What Works Clearinghouse, 2022). Researchers also conducted a cost-effectiveness analysis relating each statistically favorable effect to *Mission.io*'s per-student cost.

Research Findings	
	Students who used <i>Mission.io</i> demonstrated improved content knowledge compared to non-users (<i>Hedges' g</i> = 0.335, <i>p</i> < .001). This large effect was rated Highest Value for cost-effectiveness.
	Students who used <i>Mission.io</i> demonstrated improved writing performance compared to non-users (<i>Hedges' g</i> = 0.269, <i>p</i> = .015). This large effect was rated Highest Value for cost-effectiveness.
	Students who used <i>Mission.io</i> demonstrated improved behavioral engagement compared to non-users (<i>Hedges' g</i> = 0.173, <i>p</i> = .009). This medium effect was rated High Value for cost-effectiveness.
	Students who used <i>Mission.io</i> demonstrated higher behavioral disaffection compared to non-users (reverse-coded; <i>Hedges' g</i> = -0.560, <i>p</i> < .001).

Note. Behavioral disaffection is reverse-coded so that higher scores indicate less disaffection (a more favorable outcome). The negative effect size means that *Mission.io* users reported more behavioral disaffection than non-users.

Conclusions

Given the positive finding on behavioral engagement¹, this study provides results to satisfy ESSA evidence requirements for Level II (Moderate Evidence) regarding behavioral engagement for fifth grade students.

¹ Behavioral engagement was measured with a valid, independent measure, and therefore meets all ESSA Level II criteria. Content knowledge and writing performance, which showed the largest favorable effects, were measured with internally developed instruments and accordingly qualify under ESSA Level III (Promising Evidence).

TABLE OF CONTENTS

Introduction	5
Study Design and Methods.....	6
Implementation	11
Usability	12
Student Outcomes	15
Conclusions.....	19
References	21
Appendix A. Additional Findings by Implementation Context.....	22
Appendix B. Student Survey Instrument	24

INTRODUCTION

Mission.io is a collaborative, simulation-based learning platform that engages whole classrooms in narrative-driven, standards-based Missions where students work together to solve problems, write, design, and make decisions across academic content areas.

As part of their ongoing efforts to demonstrate the effectiveness of their solution, *Mission.io* contracted with Instructure to examine the impact of *Mission.io* on fifth grade students' academic, engagement, and disaffection outcomes. Using the Every Student Succeeds Act (ESSA) standards as guidance in developing a study design, this report align with Level II requirements (Moderate Evidence). The following research questions guided this study:

Implementation

1. To what extent are fifth grade students using *Mission.io*?
 - a. Specifically, how many *Mission.io* sessions are fifth grade students completing?

Usability

2. What are fifth grade students' perceptions of the usability of the product?

Student Outcomes

3. What is the relationship between the number of *Mission.io* sessions completed and fifth grade students' outcomes in:
 - a. Engagement
 - b. Writing self-efficacy
 - c. Content knowledge
 - d. Writing performance
4. How does the use of *Mission.io* impact fifth grade students in the following outcome areas when compared to those who did not use it?
 - a. Engagement
 - b. Writing self-efficacy
 - c. Content knowledge
 - d. Writing performance
5. What is the value of *Mission.io*, relating its measured effects to its per-student cost, including direct and indirect costs?

This report details the study design and methods, implementation, findings, and conclusions.



STUDY DESIGN AND METHODS

This section of the report briefly describes the study participants, measures, and analysis methods.

Study Design

This study used a quasi-experimental design to align with ESSA Level II evidence standards. It included fifth grade students who participated in *Mission.io* during the 2025–26 school year and a comparison sample of fifth grade students who did not use *Mission.io*. Outcome-specific propensity-score weighting was used to account for potential pre-existing differences between groups.

The primary impact analysis compared the full treatment group to the full comparison group. *Mission.io* was implemented in two contexts: a science, technology, engineering, and mathematics (STEM)-supplement context and a primary-classroom context. In the Classroom model, *Mission.io* was implemented into normal instructional routines by the core fifth grade classroom teacher (e.g., a teacher runs Missions with her 30-student fifth grade class). In the STEM model, *Mission.io* was implemented by a STEM teacher in a STEM/specialty context (e.g., each fifth grade class rotates to the STEM teacher, who runs *Mission.io* Missions with each class). STEM classrooms used slightly different rotation models; some STEM teachers saw each class for 45 minutes weekly, while others saw each class for 35 minutes every other week. Breakdowns comparing each context to non-users are reported in Appendix A.

Setting and Participants

The study was conducted in a large suburban public school district in the Mountain West region of the United States. According to data from the National Center for Education Statistics (NCES), the district serves approximately 59,000 students across more than 70 schools in a large-suburban locale, with a student body that is approximately 70% White and 20% Hispanic, with a student/teacher ratio of about 22:1.

This study included 1,039 fifth grade students across 53 classrooms, with the treatment group totaling 582 students across 28 classrooms who used *Mission.io* and the comparison group totaling 457 students across 25 classrooms who did not use the platform. By implementation context, the 53 classrooms comprised 16 STEM treatment, 18 STEM comparison, 12 classroom treatment, and seven classroom comparison classrooms. Table 1 shows the demographic breakdown of the study's participants.

Table 1. Overview of Study Participants

Demographic	Treatment Students	Comparison Students
Race/Ethnicity		
American Indian/Alaska Native	1%	1%
Asian	5%	3%
Black or African American	3%	2%
Hispanic	24%	19%
Native Hawaiian/Pacific Islander	2%	4%
White	92%	94%
Gender		
Female	49%	50%
Male	51%	50%
Program Status		
English Language Learner (ELL)	11%	4%
Free or Reduced Price Lunch (FRL)	28%	14%
Special Education (SPED)	18%	16%

Note. Race/ethnicity categories were reported as select-all-that-apply indicators; because students could select more than one category, percentages do not sum to 100%.

Measures

Student survey. The student survey included items on perceptions of *Mission.io* usability, student engagement, and writing self-efficacy. Students' perceptions of the usability of the product were captured using a six-item ease-of-use scale (1 = Very Hard to 5 = Very Easy), two single items on enjoyment and perceived learning, and open-ended questions analyzed qualitatively. Engagement was measured with the Engagement versus Disaffection with Learning (EvsD) student report (Skinner, Kindermann, & Furrer, 2009), which includes four subscales: behavioral engagement ($\alpha = .72$), emotional engagement ($\alpha = .78$), behavioral disaffection ($\alpha = .82$), and emotional disaffection ($\alpha = .85$). All EvsD items used a 4-point response scale (1 = Not at all true to 4 = Very True). Behavioral engagement and emotional engagement each contained five items, behavioral disaffection contained five items, and emotional disaffection contained 12 items. Behavioral disaffection and emotional disaffection items were reverse-coded so that higher scores indicate less disaffection (a more favorable outcome). The four subscales could not be combined and were analyzed separately. Writing self-efficacy was measured with the Self-Efficacy for Writing Scale (SEWS; Bruning, Dempsey, Kauffman, McKim, & Zumbrunn, 2013), whose subscales include ideation ($\alpha = .90$), conventions ($\alpha = .85$), and self-regulation ($\alpha = .88$); each subscale ranges 0–100 and the three were averaged into a single composite score. The survey instrument appears in Appendix B.

Content knowledge. *Mission.io* and Instructure researchers co-created a 15-item multiple-choice assessment to evaluate the intervention’s impact on academic achievement in science (assessing Utah SEEd Standards Strands 1 and 2) and social studies (assessing Utah Social Studies Core Standards 5.1.3, 5.2.1, 5.2.2, and 5.2.3). The assessment includes seven science questions and eight social studies questions, graded as one point per question with results reported as percent correct. The exact same assessment was administered at the beginning and end of the study.

Writing assessment. To evaluate students' writing proficiency, students completed two science-themed writing prompts, one in November 2025 and one in April 2026. Each prompt was accompanied by two short source texts that students read before crafting an informative/explanatory paragraph response. A team of trained, independent scorers assessed the submissions using the Utah RISE Grade 5 Writing Rubric (1–6). The first prompt concerned animal adaptations to winter and the second concerned how roller coasters work safely. The two prompts followed the same structure and were parallel in design, though the prompt content differed across the two time points.

Cost. Researchers gathered per-student pricing and training-cost information from *Mission.io* to estimate the per-student cost of implementation, including direct and indirect costs, for the cost-effectiveness analysis.

Procedures. Participating classrooms were recruited within the district across the two implementation contexts. Parental and student opt-out forms were used to protect participant rights. Each student received a unique ID to keep responses de-identified throughout data collection, which was used to link platform usage data and district-provided demographic information with pre- and post-test responses. The pretest survey and assessments were administered at the beginning of the implementation window, and the identical posttest was administered at the end of the window. Measures were largely consistent across time points, with some differences by instrument. The content knowledge assessment was identical at pretest and posttest. The student survey was identical across time points for comparison students; treatment students completed the same survey at both time points but received additional usability and feasibility items at posttest. The two writing prompts followed the same structure and were parallel in design, though the prompt content differed across time points.

- **Preparation (May–September 2025):** Researchers finalized all study instruments and submitted an application to the Institutional Review Board (IRB). By September 2025, the school district signed a data-sharing agreement with Instructure.
- **Participant Onboarding (October 2025):** Instructure researchers conducted a 30-minute virtual orientation via Zoom to introduce the pilot study to participating school district personnel. During this orientation, researchers also instructed teachers in the comparison group on the importance of ensuring their students did not use *Mission.io* during the study period.
- **Initial Data Collection (November 2025):** Using Qualtrics, Instructure researchers distributed teacher consent, parental permission opt-out forms, and child assent forms, along with the pre-survey, content knowledge assessment, and first writing prompt to teachers for distribution to students. Parental permission was obtained using a digital opt-out form; permission was implied unless the parent or guardian checked a box on the online form to decline participation. Child assent was obtained electronically before any research activities began. The pre-test data collection window was extended for STEM classrooms, since those teachers see students less frequently than classroom teachers.

- Ongoing Data Collection (December 2025–April 2026): *Mission.io* team members encouraged treatment teachers to use the *Mission.io* product regularly with their students. *Mission.io* monitored student usage data throughout this period. Students whose parents had opted them out did not participate in any data collection activities, though treatment-group students could still participate in *Mission.io* learning experiences as part of regular classroom instruction.
- Final Data Collection (April 2026): Students completed the post-surveys, the second writing prompt, and the content knowledge assessment. Researchers also collected SIS/demographic data from the school district and *Mission.io* usage data for consenting participants. Cost information (per student, direct and indirect costs) was obtained from *Mission.io*.
- Data Analysis and Reporting (May–June 2026): Instructure researchers conducted data analysis in May and June 2026. Comprehensive reporting of the study findings was completed by June 2026.
- Compensation: Treatment-group teachers received a \$300 stipend and comparison-group teachers received a \$100 stipend, delivered via e-gift card within 4 weeks of the study's conclusion. Students did not receive any payment or compensation for participation.
- Ethical Oversight: The study was reviewed and approved by the BRANY IRB (Protocol #25-235, approved 10/09/2025) and by the school district's Research Review Committee (9/23/2025). All data security and confidentiality procedures complied with the Family Educational Rights and Privacy Act (FERPA) and Utah Student Data Privacy and Data Protection requirements.

Data Analysis

Mission.io uploaded data from the 2025–26 school year through a secure file transfer protocol. Researchers characterized usage using descriptive statistics. To estimate impacts, researchers used regression models with outcome-specific propensity-score weighting, retained the matched pretest as a covariate, and applied cluster-robust standard errors at the classroom level. Clustering of students within classrooms was accounted for using cluster-robust (Huber–White) standard errors at the classroom level within the weighted regression models, rather than through multilevel modeling. Researchers calculated standardized effect sizes (*Hedges' g* with a small-sample correction) to determine the magnitude of changes in treatment students' outcomes. Researchers also conducted a cost-effectiveness analysis, relating each statistically favorable effect to *Mission.io*'s per-student cost using the framework described by Kraft (2020).

Propensity Score Weighting and Baseline Equivalence

To balance the treatment and comparison groups, researchers applied outcome-specific propensity-score weighting. Because weights were estimated separately for each outcome, analytic sample sizes differ slightly across outcomes; this approach retained as many students as possible in each analysis. To adhere to ESSA Level II standards, researchers assessed the baseline equivalence of all outcome measures between the treatment and comparison groups. Baseline equivalence on pretest outcomes ranged from a *Hedges' g* of -0.079 to 0.137 , within the range requiring statistical adjustment but not exceeding the tolerance for inclusion.

Demographic imbalance on ELL status ($h = 0.261$) and FRL status ($h = 0.352$) was addressed through the propensity-score weighting approach in all subsequent impact analyses. All other demographic characteristics were within the acceptable baseline range. Full baseline equivalence results appear in Table 2.

Table 2. Baseline Equivalence for Pretest Outcomes Measures (Full Sample)

Pretest Outcome	Treatment Mean	Comparison Mean	Hedges' g
Behavioral Engagement	3.38	3.38	-0.005
Emotional Engagement	3.15	3.07	0.137
Behavioral Disaffection (rev)	3.33	3.36	-0.057
Emotional Disaffection (rev)	3.24	3.24	-0.007
Writing Self-Efficacy	67.86	69.12	-0.058
Content Knowledge	0.38	0.38	-0.023
Writing Performance	2.83	2.93	-0.079

Note. Demographics showed baseline imbalance on ELL ($h = 0.261$) and FRL ($h = 0.352$), both addressed through propensity-score weighting. All other demographic covariates (gender, race/ethnicity indicators, and SPED status) fell within acceptable baseline equivalence thresholds and did not require additional adjustment.

IMPLEMENTATION

To better understand the extent of student engagement with the platform, this section examines how fifth grade students used *Mission.io*. A single *Mission.io* experience typically runs between 40 and 60 minutes.

To what extent are fifth grade students using *Mission.io*?

Treatment students completed an average of 9.4 *Mission.io* sessions (median = 9, range 1–20) over the implementation window, with most students completing between 5 and 13 sessions. Table 3 shows session completion among treatment students.

Table 3. *Mission.io* Session Completion Among Treatment Students

Usage Metric	<i>N</i>	Mean	<i>SD</i>	Median
All Treatment	580	9.42	4.70	9

Session completion was relatively consistent across demographic subgroups. No large equity gaps emerged in who engaged with the platform, though ELL students showed slightly lower completion. Table 4 shows mean session completion by subgroup.

Table 4. *Mission.io* Session Completion by Demographic Subgroup (Treatment Students)

Subgroup	In-group <i>N</i>	In-group Mean	Comparison Mean	Gap
English Language Learner	63	8.0	9.6	1.6
Free or Reduced-Price Lunch	162	8.8	9.7	0.9
Special Education	101	8.9	9.5	0.6
Hispanic	142	8.8	9.6	0.8

Note. In-group mean reflects students in the named subgroup; comparison mean reflects treatment students not in that subgroup (e.g., non-ELL students). Gap is the difference between the two. Subgroups are not mutually exclusive.

USABILITY

This section examines how fifth grade students perceived the usability of *Mission.io*, drawing on a six-item ease-of-use survey, two single-item enjoyment and perceived-learning measures, and an analysis of open-ended responses.

What are fifth grade students' perceptions of the usability of the product?

Table 5. Usability Survey Item Means

Item	Mean	SD	Reading
How easy is it to use <i>Mission.io</i> ?	4.01	1.03	Easy
How easy is it to use the tools to write during <i>Mission.io</i> ?	3.71	1.10	Easy
How easy is it to use the tools to draw or design during <i>Mission.io</i> ?	3.48	1.18	Okay
How easy is it to make new things and share your ideas during <i>Mission.io</i> ?	3.60	1.13	Easy
How easy is it to work together with your classmates during <i>Mission.io</i> ?	3.66	1.21	Easy
How easy is it to see how you are doing on your work during <i>Mission.io</i> ?	3.79	1.12	Easy

Subgroup patterns. ELL students rated usability notably lower ($M = 3.28$) than non-ELL students ($M = 3.76$). FRL students rated lower ($M = 3.55$) than non-FRL students ($M = 3.77$). Male students rated usability higher ($M = 3.87$) than female students ($M = 3.54$).

Student enjoyment and perceived learning. For ease of interpretation, both items were reverse coded that higher values indicate more positive perceptions. Student responses to “How much do you like using *Mission.io* in class?” (reverse coded, $M = 3.89$ on a 1–5 scale, where 5 = “I like it a lot”) and “How much does *Mission.io* help you learn?” (reverse coded, $M = 3.08$ on a 1–4 scale, where 4 = “It helps me a lot”) both fell near the most-positive end of each scale, indicating broadly favorable student perception.

Qualitative coding. The qualitative review analyzed two open-ended fields, “What do you like the most about using *Mission.io*?” and “Is there anything else you want to tell us about using *Mission.io*?”, across 1,039 student records, yielding a potential 2,078 responses. Of these, 1,030 were valid non-blank responses and 1,048 were blank. Coded themes are summarized below, with representative student quotes.

Engagement and Enjoyment [281 mentions]

The most prominent theme was general positive sentiment. Students found the platform highly engaging, a strong indicator of good emotional usability.

"It's fun because you get to connect with classmates and have fun with them while you're also focusing on work."

"I like the thrill and adventure"

"I like how Mission.io makes learning fun and engaging."

Teamwork and Collaboration [194 mentions]

Students frequently cited the collaborative aspects as their favorite part of the platform's design.

"You get to solve problems and you work together with your teams. Team WORK!"

"What I like most about mission.io is having different people to work with."

"I like the new experiences and talking with classmates."

Story and Content [158 mentions]

Students were responsive to the narrative, themes, and visual graphics of the missions.

"The missions are always different and require us to think outside the box."

"The missions are interesting"

"I like how in Mission.io it comes up with new and original ideas and it helps me learn new things."

Learning [150 mentions]

Students recognized the platform as an educational tool and appreciated the integration of learning into the gameplay loop.

"I like how you learn but it is still a lot of fun."

"It is really fun and is a fun way to review and to learn."

"I learn new things and love working together."

Game Mechanics [85 mentions]

A subset of responses focused on driving and controlling the spaceship, highlighted as a peak interactive moment.

"I love when we get to dock the ship and control the ship."

"It is super fun and I love driving the ship and scanning the areas."

"That we can work together to fly the ship"

Negative Perceptions and Frustrations [70 mentions]

Students mentioned issues with peers deleting their work, glitches, certain writing tasks taking too long, or controls feeling difficult.

"Writing 5 paragraphs are annoying"

"Sometimes we can't hear the audio, or it will glitch then kick us out then we have to rejoin."

"When I try to do something it gets deleted..."

Specific Usability and UI Callouts [22 mentions]

A smaller subset specifically praised or critiqued the user interface, the ease of logging in, and the creative tools.

"It is pretty easy to log on and play it through."

"It's fun and it's easy to use."

"I like how Mission.io gives me many tools while doing lessons."

"It can be hard to use."

"I like driving the spaceship but it is kind of hard for me to read the directions and know what to do"

Likes Social Studies [8 mentions]

A recurring subject-specific theme was enthusiasm for the social studies and history content embedded in the missions. Students singled out the history-themed missions as a favorite part of the experience. No responses expressed dislike of the social studies content, so only a positive social studies code was applied.

"The American history ones."

"I like all the history things and the space theme."

"That we learn about historical and other types of stuff."

STUDENT OUTCOMES

To address student outcomes, researchers completed two types of analyses. Dose-response analyses among treatment students used multilevel mixed models with random intercepts at the classroom level to account for within-classroom clustering. Comparative analyses between the full treatment and comparison groups used linear regression with outcome-specific propensity-score weighting and cluster-robust standard errors at the classroom level. Both analyses retained the matched pretest as a covariate and controlled for demographic characteristics (gender, race/ethnicity, ELL status, FRL status, and SPED status). Breakdowns by implementation context appear in Appendix A.

What is the relationship between the number of sessions completed and fifth grade students' outcomes?

Among treatment students, researchers estimated how outcomes shifted with each additional *Mission.io* session completed, controlling for the outcome-specific pretest and demographics. Demographic covariates included gender, five race/ethnicity indicators, ELL status, FRL status, and SPED status. Table 6 shows the dose-response results.

Table 6. Dose-response: Change Per Session Completed

Outcome	Change per session	p-value
Behavioral Engagement (1–4 scale)	–0.0002	.966
Emotional Engagement (1–4 scale)	–0.002	.775
Behavioral Disaffection rev (1–4 scale)	0.013	.061
Emotional Disaffection rev (1–4 scale)	0.001	.796
Writing Self-Efficacy (0–100 scale)	–0.08	.687
Content Knowledge (proportion correct)	0.0094	<.001*
Writing Performance (1–6 rubric)	0.035	.049*

Note. Statistically significant findings ($p < .05$) are marked with an asterisk (*).

For content knowledge, each additional session was associated with a gain of 0.0094 in proportion correct, equivalent to a 0.94 percentage-point increase. Across the observed dose range, students who completed all 20 sessions scored approximately 17.9 percentage points higher on content knowledge than students who completed just one session. This was the strongest dose-response signal in the study ($p < .001$).

For the writing rubric, each additional session was associated with a 0.035-point gain on the writing rubric ($p = .049$). Across the full observed dose range, students who completed all 20 sessions scored approximately 0.67 points higher on the 1–6 rubric than students who completed just one session.

No detectable dose-response relationships emerged for self-efficacy or engagement or disaffection.

How does the use of *Mission.io* impact fifth grade students compared to those who did not use it?

In the primary impact analysis comparing the treatment group to the comparison group, *Mission.io* produced significant favorable effects on content knowledge (*Hedges' g* = 0.335, $p < .001$), writing performance (*Hedges' g* = 0.269, $p = .015$), and behavioral engagement (*Hedges' g* = 0.173, $p = .009$). The behavioral disaffection (reverse-coded) negative result is also reported in Table 7. The remaining outcomes did not reach statistical significance.

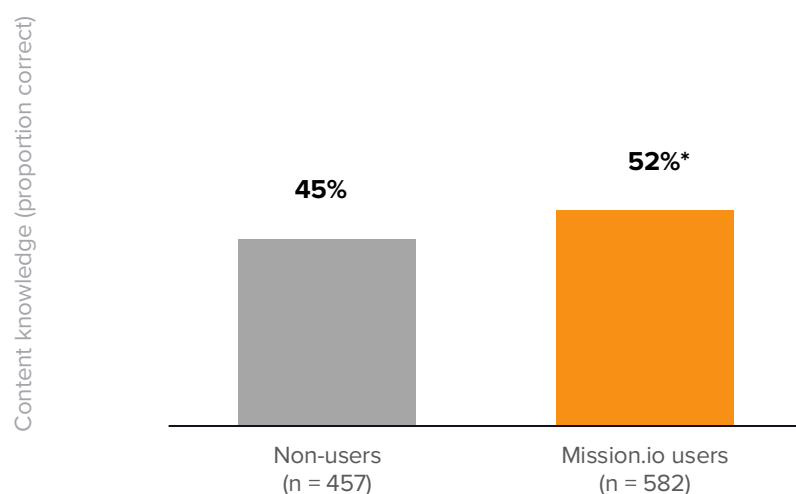
Table 7. Primary Impact: Full Treatment vs. Full Comparison

Outcome	Adj. MD	<i>Hedges' g</i>	<i>p</i> -value
Behavioral Engagement	0.092	0.173**	.009
Emotional Engagement	0.085	0.124	.123
Behavioral Disaffection (rev)	-0.371	-0.560***	<.001
Emotional Disaffection (rev)	0.037	0.057	.291
Writing Self-Efficacy	0.742	0.034	.588
Content Knowledge	0.065	0.335***	<.001
Writing Performance	0.303	0.269*	.015

Note. Statistically significant findings ($p < .05$) are marked with an asterisk (*).

Figure 1. Posttest Content Knowledge for *Mission.io* Users and Non-users.

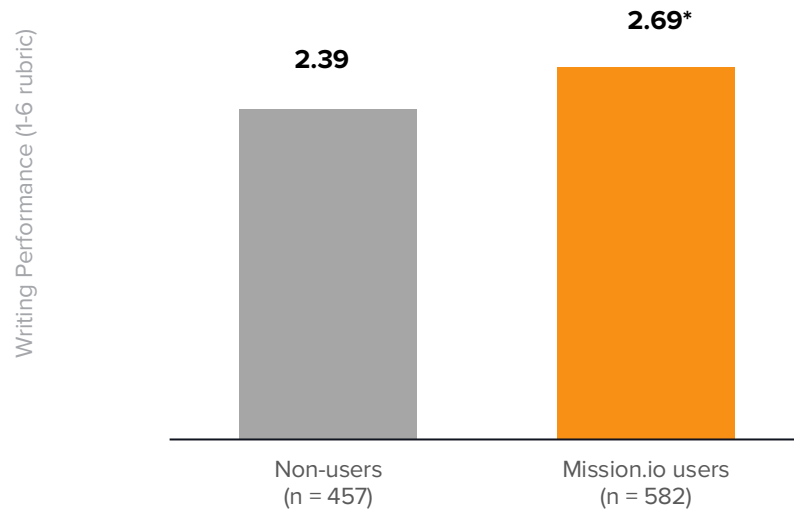
Mission.io users scored higher on content knowledge compared to non-users.



Note. Statistically significant findings ($p < .05$) are marked with an asterisk (*). Bars show marginal means. Content knowledge is reported as proportion correct (*Hedges' g* = 0.335).

Figure 2. Posttest Writing Performance for *Mission.io* Users and Non-users.

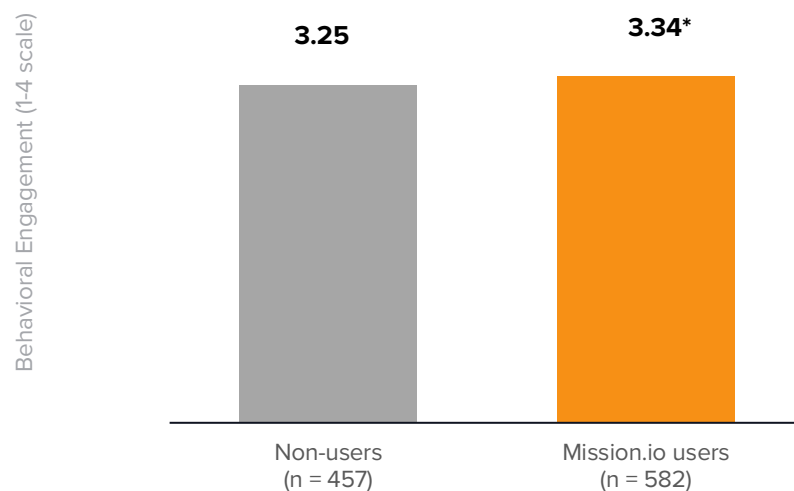
***Mission.io* users scored higher on writing performance than non-users.**



Note. Statistically significant findings ($p < .05$) are marked with an asterisk (*). Bars show observed posttest means on the 1–6 writing rubric (Hedges' $g = 0.269$).

Figure 3. Posttest Behavioral Engagement for *Mission.io* Users and Non-users

***Mission.io* users showed higher behavioral engagement than non-users**



Note. Statistically significant findings ($p < .05$) are marked with an asterisk (*). Bars show observed posttest means on the 1–4 behavioral engagement scale (Hedges' $g = 0.173$).

What is the per student cost of using *Mission.io*?

Mission.io's cost per student, including direct and indirect costs, is under \$500. *Mission.io* is licensed at an average cost of \$3.00 per student based on district-level pricing; for the district in this study, pricing would fall between \$2.14 and \$2.85 per student annually given its enrollment size. Pricing ranges from \$2.14 to \$6.00 per student per year depending on district size and contract length. Standard training is included (e.g., online training modules, virtual kickoff, quarterly webinars), with additional professional development support available for purchase (e.g., custom virtual workshops, on-site training). Following Kraft (2020), each statistically favorable effect was related to this per-student cost to produce a cost-effectiveness rating. Table 8 summarizes the cost-effectiveness rating for each statistically favorable effect in the primary analysis.

Table 8. *Mission.io* Cost-Effectiveness Ratings by Outcome

Outcome	Hedges' <i>g</i>	Effect Size Bin	Cost/Student	Cost-effectiveness Rating
Content Knowledge	0.335***	Large ($\geq .20$)	< \$500 (Low)	Highest Value
Writing Performance	0.269*	Large ($\geq .20$)	< \$500 (Low)	Highest Value
Behavioral Engagement	0.173**	Medium ($.05 - < .20$)	< \$500 (Low)	High Value

CONCLUSIONS

Conclusions, Unanticipated Outcomes, Barriers, and Recommendations

The research results demonstrate that *Mission.io* produced meaningful, positive effects on student learning and engagement. Students who used the platform showed significant gains in content knowledge (*Hedges' g* = 0.335, $p < .001$) and writing performance (*Hedges' g* = 0.269, $p = .015$), with behavioral engagement also improving favorably (*Hedges' g* = 0.173, $p = .009$). These results suggest the product successfully delivers on its promise to improve academic outcomes in school settings. The cost-effectiveness analysis further strengthened this conclusion: both content knowledge and writing performance effects were rated “Highest Value.”

One unanticipated finding was that the dose-response analysis confirmed that each additional mission completed was associated with meaningful gains in content knowledge and writing performance, providing evidence that continued engagement with the platform accelerates academic benefit.

On the behavioral disaffection subscale (reverse-coded), *Mission.io* users reported more disaffection than non-users (*Hedges' g* = -0.560, $p < .001$). A negative effect indicates that users endorsed disaffection-related items, such as going through the motions or doing just enough to get by, at higher rates than non-users. This pattern sits alongside the favorable behavioral engagement finding, which suggests that students were not less engaged overall but may have experienced moments of frustration or disengagement during specific tasks. Student qualitative feedback offers a plausible explanation that the most common frustrations involved writing tasks that felt too long, peers deleting shared work, and occasional technical glitches, each of which could prompt students to disengage in the moment even while remaining behaviorally engaged across the experience as a whole.

The study also revealed areas for continued development. ELL students rated the platform's usability notably lower than their non-ELL peers ($M = 3.28$ vs. 3.76), and a Spanish immersion classroom requested a Spanish version of the survey instruments, which was not available at the time of the study. This points to an opportunity to invest in multilingual support across both the platform and research instruments to better serve diverse learners.

Student qualitative feedback also surfaced frustrations around peers deleting work, occasional glitches, and writing tasks that felt too long, all of which are areas the product team can address in future iterations. Teachers should also establish clear norms around collaborative workspaces before students begin missions to reduce disruptions.

With respect to ESSA evidence tiers, the behavioral engagement finding, measured with the valid and independent EvsD scale, satisfies ESSA Level II (Moderate Evidence) requirements; the content knowledge and writing performance findings, which used internally developed measures, qualify under ESSA Level III (Promising Evidence). Although the results indicate one unfavorable finding (Behavioral Disaffection), the favorable findings outweigh the unfavorable findings in this report.

Potential Problem Areas and How They Were Addressed

Several logistical and technical challenges arose during the study, each of which the team worked to address. First, the pre-test data collection window had to be extended for STEM classrooms, since those teachers see students less frequently than classroom teachers do. This

extension may have resulted in a slightly later implementation start for some STEM classrooms relative to their classroom counterparts, though every effort was made to minimize that gap.

Second, the study used parental and student opt-out forms to protect participant rights. Some students accidentally opted out during the pre-test phase, which required the *Mission.io* team to work directly with teachers to verify each case and confirm whether those responses could legitimately be included in the study.

Third, each student received a unique ID to keep responses de-identified throughout data collection. Some students did not enter their IDs correctly, making it impossible to link their platform usage data to their pre- and post-test results. As a result, 93 students had to be excluded from the pre-test sample and 41 from the post-test sample.

Fourth, one teacher reported a technical issue during survey administration in which students' writing submitted prematurely when they pressed enter, and at least one student's text was displaying right to left. These issues disrupted the experience for those students. Additionally, a Spanish immersion class requested a Spanish version of the survey instruments, which wasn't available for this study. That gap is a clear area to address before future research cycles.

Finally, five classrooms dropped from the study between pre- and post-test, and 19 students moved or changed classes during the implementation window and had to be removed from the analyses. While attrition of this kind is common in school-based research, it does reduce sample size and limits the conclusions that can be drawn from those classrooms.



REFERENCES

- Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology, 105*(1), 25–38.
- Every Student Succeeds Act, Pub. L. No. 114-95 (2015).
<https://www.govinfo.gov/app/details/PLAW-114publ95>.
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253.
- Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children’s behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493–525.
- What Works Clearinghouse (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). This report is available on the What Works Clearinghouse website at <https://ies.ed.gov/ncee/wwc/Handbooks>.

APPENDIX A. ADDITIONAL FINDINGS BY IMPLEMENTATION CONTEXT

This appendix reports findings broken out by implementation context, presented together (Combined, STEM, Classroom). *Mission.io* was implemented in two contexts. In the Classroom model, *Mission.io* was implemented into normal instructional routines by the core fifth grade classroom teacher. In the STEM model, *Mission.io* was implemented by a STEM teacher in a STEM/specialty rotation context; some STEM teachers saw each class for 45 minutes weekly, while others saw each class for 35 minutes every other week. The STEM-context comparison included 631 students (312 treatment, 319 comparison) and the classroom-context comparison included 408 students (270 treatment, 138 comparison).

Implementation

Table A1. *Mission.io* Session Completion by Implementation Context

Group	N	Mean	SD	Median	Range
STEM Treatment	310	6.32	2.85	5	1–12
Classroom Treatment	270	12.97	3.81	13	5–20

Student Outcomes

Dose-response by context. Among treatment students in each context, content knowledge and writing performance both showed significant dose-response relationships, while engagement, disaffection, and self-efficacy did not yield any significant findings. The STEM coefficient on writing performance (0.10 per session) was larger than the corresponding classroom coefficient (0.07). The same pattern held for content knowledge, where the STEM coefficient (0.016 per session) was also larger than the classroom coefficient (0.014)

Table A2. Dose-response by Session, by Implementation Context

Outcome	STEM Tx coef (<i>p</i>)	Classroom Tx coef (<i>p</i>)
Behavioral Engagement	–0.0092 (.526)	0.0131 (.203)
Emotional Engagement	0.0133 (.567)	0.0159 (.231)
Behavioral Disaffection (rev)	0.0140 (.380)	0.0156 (.235)
Emotional Disaffection (rev)	0.0155 (.256)	–0.0055 (.682)
Writing Self-Efficacy	–0.537 (.322)	0.595 (.226)
Content Knowledge	0.0160 (.034*)	0.0135 (.004*)
Writing Performance (1–6)	0.1005 (.013*)	0.0666 (.022*)

Note. **p* < .05.

Impact vs. non-users by context. Compared to non-users, STEM-context users showed significant favorable effects on content knowledge (*Hedges' g* = 0.355**), writing performance (*Hedges' g* = 0.422**), behavioral engagement (*Hedges' g* = 0.167*), and emotional engagement (*Hedges' g* = 0.221*). Classroom-context users showed no significant favorable effects. Behavioral disaffection (reverse-coded) showed a significant unfavorable effect in both contexts (STEM *Hedges' g* = -0.475***; Classroom *Hedges' g* = -0.753***).

Table A3. Impact vs. Non-Users (*Hedges' g*) by Implementation Context

Outcome	STEM <i>Hedges' g</i>	Class <i>Hedges' g</i>
Behavioral Engagement	0.167*	0.140
Emotional Engagement	0.221*	-0.060
Behavioral Disaffection (rev)	-0.475*	-0.753*
Emotional Disaffection (rev)	0.097	-0.016
Writing Self-Efficacy	0.099	-0.066
Content Knowledge	0.355*	0.124
Writing Performance	0.422*	0.140

Note. Combined *n* = 1,039; STEM *n* = 631; Classroom *n* = 408. Green = significant favorable; red = significant unfavorable. **p* < .05.

APPENDIX B. STUDENT SURVEY INSTRUMENT

The student survey administered for this study is reproduced below. The engagement and writing self-efficacy items were administered at both pretest and posttest. The usability and feasibility items (items 45–54), including the single enjoyment and perceived-learning items, were administered at posttest only and to treatment students. The content knowledge and writing assessments were administered at both pretest and posttest.

Introduction and Assent

Mission.io is working with Instructure to do a research study. We want to understand how *Mission.io* helps you feel about doing your schoolwork and how you feel about writing. We are asking you to help by answering some questions for our research study. You do not have to answer any questions that you do not want to and your answers will be kept private. If you want to help with our research study, please check the box below.

- ☐ Yes, I want to be in the study and answer the questions.
- ☐ No, I don't want to be in the study.

Information

1. What is your study ID number?

Engagement (4-point scale: Not at all true / A little true / Mostly true / Very true)

2. I try hard to do well in school.
3. In class, I work as hard as I can.
4. When I'm in class, I participate in class discussions.
5. I pay attention in class.
6. When I'm in class, I listen very carefully.
7. When I'm in class, I feel good.
8. When we work on something in class, I feel interested.
9. Class is fun.
10. I enjoy learning new things in class.
11. When we work on something in class, I get involved.
12. When I'm in class, I just act like I'm working.
13. I don't try very hard at school.
14. In class, I do just enough to get by.
15. When I'm in class, I think about other things.
16. When I'm in class, my mind wanders.
17. When we work on something in class, I feel bored.
18. When I'm doing work in class, I feel bored.
19. When my teacher first explains new material, I feel bored.
20. When I'm in class, I feel worried.
21. When we start something new in class, I feel nervous.
22. When I get stuck on a problem, I feel worried.
23. When we work on something in class, I feel discouraged.
24. Class is not all that fun for me.
25. When I'm in class, I feel bad.

26. When I'm working on my classwork, I feel mad.
27. When I get stuck on a problem, it really bothers me.
28. When I can't answer a question, I feel frustrated.

Writing Self-Efficacy (0 = not confident at all to 100 = very confident)

29. I can think of many ideas for my writing.
30. I can put my ideas into writing.
31. I can think of many words to describe my ideas.
32. I can think of a lot of original ideas.
33. I know exactly where to place my ideas in my writing.
34. I can spell my words correctly.
35. I can write complete sentences.
36. I can punctuate my sentences correctly.
37. I can write grammatically correct sentences.
38. I can begin my paragraphs in the right spots.
39. I can focus on my writing for at least one hour.
40. I can avoid distractions while I write.
41. I can start writing assignments quickly.
42. I can control my frustration when I write.
43. I can think of my writing goals before I write.
44. I can keep writing even when it's difficult.

Usability (post-survey only) (5-point scale: Very hard / Hard / Okay / Easy / Very easy)

45. How easy is it to use *Mission.io*?
46. How easy is it to use the tools to write during *Mission.io*?
47. How easy is it to use the tools to draw or design during *Mission.io*?
48. How easy is it to make new things and share your ideas during *Mission.io*?
49. How easy is it to work together with your classmates during *Mission.io*?
50. How easy is it to see how you are doing on your work during *Mission.io*?
51. How much do you like using *Mission.io* in class? (I like it a lot / I like it / It is okay / I do not like it much / I do not like it at all)
52. What do you like the most about using *Mission.io*? (open response)
53. How much does *Mission.io* help you learn? (It helps me a lot / It helps me a little / It does not help me much / It does not help me at all)
54. Is there anything else you want to tell us about using *Mission.io*? (open response)

Content knowledge assessment

Science

1. Which of the following statements is true about plate tectonics?
 - a. Volcanoes and earthquakes are random and there is no pattern to where they form.
 - b. Plate tectonics only affects the weather, not landforms.
 - c. Volcanoes and earthquakes are most likely to form on the borders of a plate.
 - d. Volcanoes and earthquakes are most likely to form in the middle of a plate.
 - e. Earthquakes only happen in the middle of continents, not near oceans.
2. Which of the following statements is TRUE?
 - a. Most of Earth's water is found in lakes and rivers.
 - b. All water is accessible on Earth.

- c. There is more saltwater than freshwater on Earth.
 - d. There is no accessible freshwater on Earth.
 - e. Earth has equal amounts of freshwater and saltwater.
3. What's the most likely cause of the erosion shown in this image?



- a. Chemical
 - b. Vegetation
 - c. Water
 - d. Wind
 - e. Glacial
4. Sugar disappears when mixed into a glass of water. Which statement best describes what happens to the sugar?
- a. The sugar disappears when mixed with water.
 - b. The sugar transforms into water when mixed.
 - c. The sugar breaks into tiny particles.
 - d. The sugar floats on top, even though we can't see it.
 - e. The sugar melts because the water is a liquid.
5. A scientist stirs a spoonful of powder into water and watches it dissolve completely. What does this tell us about the powder?
- a. The powder is a gas.
 - b. The powder is buoyant.
 - c. The powder is metallic.
 - d. The powder is soluble.
 - e. The powder reacts with light.
6. Two clear liquids are mixed together. The temperature increases and the color changes to orange. Which statement BEST explains what happened?
- a. The liquids combined and created heat from friction while stirring.
 - b. A chemical reaction occurred, producing new substances with different properties.
 - c. The color changed because the two liquids had different densities.
 - d. The mixture warmed up because one of the liquids was already hot.
 - e. The orange color formed because air mixed in during stirring.
7. Stephen has a container of apple juice that weighs 30g. He puts the container in the freezer. How much would you expect the apple juice to weigh after it has frozen?
- a. More than 30g
 - b. Less than 30g
 - c. Exactly 30g

Social Studies

8. Which colony was founded by the Virginia Company of London for the purpose of mining gold and silver?
 - a. Maryland
 - b. Massachusetts Bay
 - c. Jamestown
 - d. Plymouth
 - e. Roanoke
9. Why did the Jamestown colony struggle? Choose TWO.
 - a. There was religious persecution that caused the population to shrink.
 - b. The colonists were unprepared and did not focus on building farms.
 - c. The colony lacked leadership and someone to be in charge.
 - d. The colonists did not have enough food and faced starvation.
 - e. The climate was too cold for anyone to survive.
10. Which of the following was NOT a reason the Patriots rebelled against the British?
 - a. British troops sometimes used force against the colonists.
 - b. The colonists were being taxed too heavily.
 - c. British soldiers were allowed to stay in colonists' homes.
 - d. The colonists had no voice in making laws or decisions.
 - e. The colonies were running out of food during the war.
11. What was the purpose of the Boston Tea Party?
 - a. To increase tea sales in the colonies.
 - b. To ask Britain for more tea.
 - c. To protest British taxes on tea.
 - d. To celebrate the King's birthday.
 - e. To show support for British trade laws.
12. Which of the following events did NOT lead to the writing of the Declaration of Independence?
 - a. Boston Massacre
 - b. Battle of Monmouth
 - c. The Intolerable Acts
 - d. Battle of Lexington and Concord
 - e. Boston Tea Party
13. What was included in the Declaration of Independence?
 - a. A plan for separating from Britain
 - b. A list of British officials
 - c. A battle plan for the Revolution
 - d. A list of grievances against the King
 - e. A request for peace negotiations with Parliament

14. Who created this political cartoon?



- a. Thomas Jefferson
 - b. Benjamin Franklin
 - c. Alexander Hamilton
 - d. George Washington
 - e. John Adams
15. Which TWO of the following statements are TRUE?
- a. France stayed neutral and refused to help the colonists during the Revolutionary War.
 - b. France helped the American colonists fight the British by sending soldiers, ships, and supplies.
 - c. Marquis de Lafayette was a British general during the war who convinced the French to form an alliance with Britain.
 - d. Benjamin Franklin and Marquis de Lafayette helped convince France to form an alliance with the American colonies.
 - e. Benjamin Franklin fought as a general on the battlefield and was responsible for the victory at the Battle of Monmouth.

Writing Assessment

1. How Do Animals Adapt to Winter?



Source Text 1: Preparing for the Cold. Winter brings freezing temperatures and less food sources, so animals must find smart ways to survive. Many animals begin preparing in the fall. For

example, squirrels gather and store nuts, seeds, and other food. They hide these in tree holes or underground so they can eat them during winter, when food is hard to find. Some animals use a different method for surviving the winter called hibernation. Hibernation is a deep sleep that helps animals save energy. Bears, bats, and groundhogs hibernate for weeks or months. Their breathing and heartbeat slow down, and their body temperature drops. They don't eat or move much during this time. Instead, they live off of the fat they stored during warmer months. This helps them stay alive without needing to search for food in the cold. Hibernation and food storage are just two survival strategies. Animals that use these methods are often less active in the winter, but they are well-prepared to make it through until spring returns.



Source Text 2: Winter Survival Strategies. To live through cold winters, animals use special behaviors and body changes called adaptations. One common adaptation is growing thick fur. As fall arrives, animals like deer, foxes, and rabbits grow warm coats to protect themselves from freezing weather. In spring, they shed this fur when it's no longer needed. Some animals also change color in winter. Arctic foxes, weasels, and hares turn white to blend in with the snow. This helps them hide from predators and sneak up on prey. This ability to blend in is called camouflage. Another helpful winter strategy is migration. Many birds, like geese and robins, fly south to warmer places where food is easier to find. Even some insects, such as monarch butterflies, travel long distances to escape the cold. These adaptations—growing fur, changing color, and migrating—help animals survive until spring. Each species uses the best method for its body and environment.

Writing Prompt: Using details from both texts, write a paragraph explaining how animals survive during winter. Describe different adaptations and strategies animals use to stay alive in cold weather. Make sure to support your explanation with evidence from both texts.

2. How Do Roller Coasters Work Safely?



Source Text 1: The Science of Thrills. Roller coasters are exciting rides that use the science of energy, gravity, and motion to give people a thrilling experience. At the beginning of a roller coaster ride, a motor or chain pulls the train of cars up a tall hill. This gives the coaster potential energy, which is stored energy that is ready to be used. The higher the hill, the more potential energy is stored.

When the coaster reaches the top and starts going down, the potential energy changes into kinetic energy—the energy of motion. The coaster speeds down the hill and through loops, curves, and drops. Gravity helps pull the coaster downward, and friction from the track and the air slows it down slightly.

Engineers and designers carefully plan every part of the ride. They use math, science, and computers to make sure the coaster has enough energy to complete the track safely. They must also make sure that the ride gives people an exciting, smooth, and safe experience. Every twist and turn is carefully calculated to thrill riders without putting them in danger.

[Image: a steel roller coaster with a loop]

Source Text 2: Staying Safe on the Ride. Safety is a top priority for roller coaster designers and theme parks. Every roller coaster is built with many features to keep riders safe from the moment the ride begins to the time it ends. These include strong seat belts, over-the-shoulder bars, and locking wheels that grip the track tightly. These safety systems keep riders secure, even when the coaster is upside down or going very fast.

Modern roller coasters also use advanced computer systems to monitor the ride in real time. These computers check the speed, brakes, and position of each train to make sure everything is working correctly. If something seems wrong, the ride can stop automatically. Each day before the ride opens, trained workers inspect the tracks, cars, and safety equipment. They make sure every part is working well. Engineers test and re-test designs using models and computer simulations before the ride is even built. Thanks to all these safety measures, millions of people ride roller coasters every year without injury, enjoying the thrills with confidence and fun.

Writing Prompt: Using information from both texts, write a paragraph explaining how roller coasters work and how they stay safe. Describe how energy helps the ride move and what safety systems are used. Support your explanation with details from both texts.



THE EDTECH COLLECTIVE

June 19, 2026

Dear Cameron Milien,

We are pleased to share that Instructure has validated the *Mission.io and Fifth Grade Learning Outcomes: An ESSA Level II Efficacy Study* as meeting the criteria for “Moderate Evidence” (ESSA Level II) under the Every Student Succeeds Act (ESSA). This designation is based on a rigorous review process grounded in multiple federal and research-based frameworks, including those from the Every Student Succeeds Act (ESSA; statute and non-regulatory guidance); Education Department General Administrative Regulations (EDGAR); Standards for Excellence in Education Research (SEER); and the What Works Clearinghouse (WWC). These frameworks ensure transparency and accessibility in evaluating research rigor and significance, helping education leaders make informed, evidence-based decisions.

The WWC version 5.0 standards offer the most comprehensive approach to validating the quality of impact studies, drawing upon the expertise of numerous research institutions and organizations, and are referenced in the US Department of Education non-regulatory ESSA guidance. Researchers can be independently certified in these standards by completing an exam. Our team includes [multiple WWC-certified reviewers](#) in the WWC Group Design (Version 4.1 and 5.0) standards, and we adhere to these standards when validating studies for the highest levels of impact (i.e., causal) evidence: Strong Evidence (ESSA Level I) and Moderate Evidence (ESSA Level II). Subsequent levels, Promising Evidence (ESSA Level III) and Demonstrates a Rationale (ESSA Level IV), are based on regulatory and non-regulatory guidelines from ESSA and the US Department of Education.

Instructure researchers use the following ESSA Rubric that evaluates five factors derived from the previously mentioned frameworks: sample, methodology, baseline equivalence, implementation and findings, and rationale. Three trained reviewers independently applied the rubric to the report, with at least one WWC-certified reviewer on every study review.

This letter and accompanying documents validate that *Mission.io* **meets “Moderate Evidence” (ESSA II) requirements** based on our review of one study across grade 5 on students’ behavioral engagement outcomes.

The ESSA rating on *Mission.io* can change as new evidence becomes available. Buyers should also consider the study population and context in their purchasing decisions.

Sincerely,

Molly Henschel, Ph.D.
Interim Director of Research, Instructure





THE EDTECH COLLECTIVE

Summary of ESSA Alignment for *Mission.io*

The following document summarizes Instructure researchers' review of *Mission.io* and provides additional information about the Every Student Succeeds Act (ESSA) review team.

Review Procedure

Instructure's research team worked with *Mission.io* to review the following study for evidence that the use of *Mission.io* impacts fifth grade students' behavioral engagement outcomes. The following document was reviewed:

- *Mission.io and Fifth Grade Learning Outcomes: An ESSA Level II Efficacy Study (Bautista, Henschel, & Allen Green, 2026)*

To determine the intervention's evidence ratings, researchers used an ESSA Validation Rubric with five factors derived from ESSA non-regulatory guidance and What Works Clearinghouse (WWC) standards and Standards for Excellence in Education Research (SEER) standards: sample, methodology, baseline equivalence, implementation and findings, and rationale. Three trained reviewers independently applied the rubric to the reports and there was excellent agreement between raters.

Study Alignment with ESSA Evidence Levels

Researchers identified the following levels of evidence:

Mission.io and Fifth Grade Learning Outcomes: An ESSA Level II Efficacy Study (2025–26) by Bautista, Henschel, and Allen Green: ESSA Level II "Moderate Evidence"

This study met 10 of 10 indicators related to ESSA's Level II "Moderate Evidence." *Mission.io* showed at least one positive, statistically significant impact on students' behavioral engagement outcomes in a well-designed and well-implemented quasi-experimental study. These results were specific to a suburban Mountain West United States context using a sample of 1,039 fifth grade learners.

In a quasi-experimental comparison of 582 *Mission.io* users and 457 non-users, students who used *Mission.io* showed a statistically significant positive effect on behavioral engagement (Hedges' $g = 0.173$, $p = .009$). In addition to the ESSA II-aligned finding, the study also found statistically significant positive effects between students who used *Mission.io* and their content knowledge (Hedges' $g = 0.335$, $p < .001$) and writing performance (Hedges' $g = 0.269$, $p = .015$). The report concludes that these findings satisfy ESSA





Level II evidence requirements, and a cost-effectiveness analysis rated the content knowledge and writing performance effects “Highest Value” at a per-student cost under \$500.

About the Review Team

Instructure researchers have over 80 years of combined experience in education research and preK–16 program evaluation. Our interdisciplinary team includes innovative researchers with diverse and advanced educational attainment (i.e., Master’s and Doctorate degrees) in Developmental Psychology, Educational Psychology, Psychometrics and Measurement, Educational Technology, Educational Research and Evaluation, Educational Research Methods, Mathematics Education, Teacher Education, and Learning Sciences.

We are certified in the What Works Clearinghouse (WWC) Group Design Standards. The following team members have WWC certification:

- Molly Henschel (Certified: Group Design (Standards Version 4.1), Group Design (Standards Version 5.0))
- Andrew Scanlan (Certified: Group Design (Standards Version 4.1), Group Design (Standards Version 5.0))
- Meetal Shah (Certified: Group Design (Standards Version 4.1), Group Design (Standards Version 5.0))
- Mary Styers (Certified: Group Design (Standards Version 4.1), Group Design (Standards Version 5.0))

Our team has extensive experience designing curriculum studies to meet WWC, Every Student Succeeds Act (ESSA), EDGAR, and SEER standards. As examples, we have collectively conducted more than 300 preK–12 curriculum ESSA evaluations for providers and more than 3,000 rapid-cycle evaluations (RCEs) in partnership with districts, including randomized control trials (RCTs), quasi-experimental designs (QEDs), and non-experimental studies (e.g., correlational design, pilot studies) across an array of content areas (e.g., science, math, literacy, early childhood, and career and technical education).

We have experience working directly with a range of small and large edtech providers and school districts on ESSA Level I–IV studies. Our team is able to meet and exceed our clients’ expectations due to our expertise in conducting quantitative and qualitative analyses (e.g., multilevel modeling, structural equation modeling, binary logistic regression, analytic induction) and in creating user-friendly reports and infographics for non-technical audiences.



Document/Product Name: _____

Date Reviewed: _____

Instructions: Read each statement below. Check the box(es) if the statement is satisfied by the document being reviewed. Boxes for each evidence level (columns) will automatically check if the statement applies to multiple levels. After all statements have been reviewed, the appropriate level of evidence is determined by examining which ESSA level has all checkboxes in its column checked. The reviewer will record the appropriate level of evidence in the final row, "ESSA Level Rating."		REQUIREMENTS			
		Demonstrates a Rationale Stage 4	Promising Evidence Stage 3	Moderate Evidence Stage 2	Strong Evidence Stage 1
SAMPLE	The study contains at least 350 students in the analysis sample.				
	There are at least two sites (i.e., local education agencies, states, locales) in the analysis sample.				
	The study specifies the setting, population, and/or subgroups of interest being served by the intervention.				
	The study has acceptable levels of attrition based on What Works Clearinghouse's (WWC) threshold for tolerable bias.				
METHODOLOGY	There are control or comparison groups (e.g., randomized controlled trial [RCT] or quasi-experimental design [QED]).				
	The participants were randomly assigned to treatment and comparison groups (e.g., RCT).				
	The outcome measure is a standardized, formal assessment administered as specified by the assessment's authors OR the outcome measure possesses all factors of a high-quality measure based on WWC and Standards for Excellence in Education Research (SEER) standards.				
	The authors use an appropriate analysis based on study design (e.g., ANCOVA for student-level assignment to conditions, MLM for classroom or school-level assignment).				
	The analysis is correlative in nature.				
	The variables of interest are clearly defined.				
	The authors include statistical controls for selection bias.				
BASELINE EQUIVALENCE	There is baseline equivalence (less than or equal to 0.05 Hedge's g effect size) on pretest measures OR authors used a matched comparison or covariates to account for pretest differences for effect sizes between 0.05 and .25				
IMPLEMENTATION & FINDINGS	The authors clearly document implementation (e.g., average dosage received).				
	There is at least one statistically significant, positive effect of the intervention on outcomes.				
RATIONALE	The authors provide a well-defined program logic model informed by research or evaluation documenting how the intervention should improve relevant outcomes.				
	There is another Rapid Cycle Evaluation (RCE) or evaluation study examining the intervention currently planned or underway, ideally a Level III study or higher.				
ESSA LEVEL RATING					

Note: If there are multiple studies on a product, ensure that the evidence is not overridden with statistically significant and negative (i.e., unfavorable) findings. Studies that do not meet Levels I or II, but statistically control for selection bias can meet Level III evidence. If the study's students and setting are not similar to local students and settings, the level of evidence rating of the intervention might change. ESSA level ratings are not static and can change as new evidence becomes available.