Study Methods for Briefs about Ecree Research Findings

This summary provides details about a study of Ecree, an automated writing feedback tool, which took place during the 2021–2022 school year. The Mathematica study team, in collaboration with Ecree, designed and conducted the study to examine teachers' and students' use of the tool during the study and the impact of Ecree on student and teacher outcomes. This summary accompanies a series of briefs summarizing results of the Ecree study for different audiences (teachers, students and families, school and district leaders, and tool designers and instructional support teams) by providing more information on the study design, sample, data sources, measures, and analytic methods used to produce the findings reported in those briefs.[1] All briefs in this series are available here: Evaluating the Development of Secondary Writing Teaching & Learning Solutions.

1 Study design

The study team designed and carried out a randomized controlled trial to examine the impact of Ecree on student outcomes.[2] The team also collected information to understand the usability, usefulness, and utilization of Ecree, and the facilitators and barriers to implementing the tool.

Sample and random assignment

The study team recruited two school districts in Alabama and North Carolina to participate in the study. Teachers who taught grades 8, 9, 10, or 11 were eligible to take part in the study and, in the summer, consented to participate. In the fall, the study team formed groups of teachers from the same district, school and grade and then randomly assigned teachers within those groups to have access to Ecree (intervention group) or to continue typical instructional practices (comparison group). Across districts, 23 teachers were assigned to the intervention group and 22 teachers were assigned to the comparison group. There were 958 students taught by teachers in the intervention group, and 870 students taught by teachers in the comparison group. Following random assignment, parents of students in participating teachers' classrooms provided their passive consent (that is, opt-out consent) and students provided their assent to participate in the study. There were 945 students taught by teachers in the intervention group, and 844 students taught by teachers in the comparison group who provided consent.

Random assignment samples

---
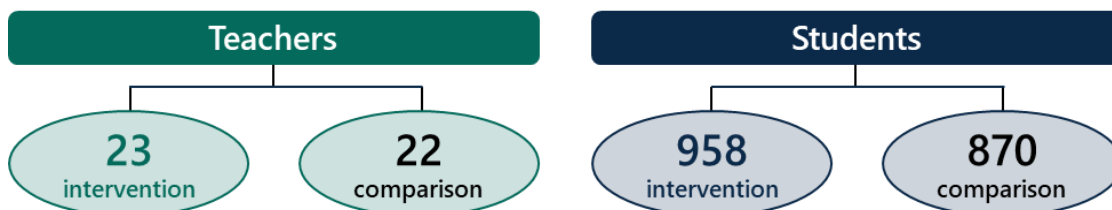
[1] This summary focuses only on methodological details related to findings reported in the series of briefs. In footnotes, we note other aspects of the study that were not included in those briefs. For additional information on the Ecree study, email tmusaddiq@mathematica-mpr.com.

[2] The study team also examined the impact of Ecree on teacher outcomes, including teachers' confidence in teaching writing and teachers' use of evidence-based practices for teaching writing.

Data collection

Data sources for the Ecree study included study-administered student and teachers surveys, study-administered writing assessments, study-administered interviews, and usage data collected from the Ecree platform. Table 1 summarizes the types of data obtained from these sources and the response rates for each.[3]

Table 1. Data collected during the study for findings presented in the briefs

Data source	Data obtained	Response rates		
		Overall	Intervention group	Comparison group
Follow-up student survey[a]	Enjoyment of writing; usability and usefulness of Ecree[a]	1,015 (56%)	547 (57%)	468 (54%)
Follow-up student writing assessment	Proficiency in argumentative writing	542 (30%)	252 (26%)	290 (33%)
Follow-up teacher survey	Usability and usefulness of Ecree	36 (80%)	18 (78%)	18 (82%)
Teacher interviews	Facilitators and barriers to implementation; classroom context	10 (n.a)[b]	7 (n.a.)[b]	3 (n.a.)[b]
Usage data collected from the Ecree platform and teacher training attendance records	Student and teacher implementation fidelity	934 students, (97%), 17 teachers (73%)	934 students (97%), 17 teachers (73%)	n.a.[c]

Note: Forty-five teachers were randomly assigned to a group (23 intervention-group and 22 comparison-group teachers). A total of 1,828 students (958 intervention-group and 870 comparison-group students) were enrolled in the classrooms of teachers who were randomly assigned. These numbers serve as denominators in response rate calculations.

[a] Sample sizes varied across survey items due to missing data among some items and skip patterns in the survey.

[b] Given the challenges presented by COVID-19 in engaging school staff during the implementation year, the study team collected interview data from a voluntary response sample of teachers that expressed a willingness to participate in interviews, so there is not a relevant denominator for calculating a response rate.

[c] Students and teachers in the comparison group did not have Ecree accounts and did not use Ecree.

n.a. = not applicable.

2 Analytic methods for estimating impacts of Ecree on student outcomes

The study team examined the impact of Ecree on several student outcomes, including students' enjoyment of writing and proficiency in argumentative writing skills. Due to low and varying response rates between the intervention and comparison groups, the study team used a statistical procedure called propensity score matching to estimate the impacts of Ecree on these outcomes.

---

[3] The study team also administered a beginning-of-year (baseline) survey to students and teachers, administered a baseline writing assessment to students, and collected administrative data on student demographics from participating school districts. We do not describe those data sources here because findings based on the samples from those data sources are not included in the briefs.

Student outcome measures[4]

To construct the student outcome measures for the study, the team collected data through both student surveys and writing assessments. These data sources incorporated the following measures:

Students' enjoyment of writing. The Liking Writing scale (LWS) measured students' enjoyment of writing on a 4-point Likert scale (strongly disagree, disagree, agree, strongly agree) across four items. The study estimated the impact of Ecree on an overall LWS score that was constructed by taking the average of the values of the four items.

Students' proficiency in argumentative writing. Writing assessments in the form of essays measured students' proficiency in argumentative writing. Human raters scored the student essays using the Smarter Balanced Argumentative Performance Task Writing Rubric which comprised three traits: (1) organization/purpose, (2) evidence/elaboration, and (3) conventions. The first two traits were rated on a 4-point scale of 1 (low) to 4 (high), and conventions was rated on a 3-point scale of 0 (low) to 2 (high). The study team estimated the impact of Ecree on each of the three traits and on the overall score that was calculated by taking the average of the first two traits plus the conventions score.

> For more information on the student measures used in this study, see the Secondary Writing Project Menu of Measures here: [Menu of Measures: Secondary Writing (mathematica.org)](mathematica.org).

Analytic samples

The final analytic sample for the survey outcome (LWS) included 433 students in the intervention group and 411 students in the comparison group. To be included in the analytic sample for the survey outcomes, students must have responded to the follow-up survey for both the LWS and the items about their confidence in writing. Students were excluded if they were the only student in their class with either score.

The final analytic sample for essay outcome included 247 students in the intervention group and 283 students in the comparison group. To be included in the analytic sample for the essay outcome, students must have completed the follow-up writing assessment and received a valid score.[5] Students were excluded if they were the only student in their class with a valid essay score.
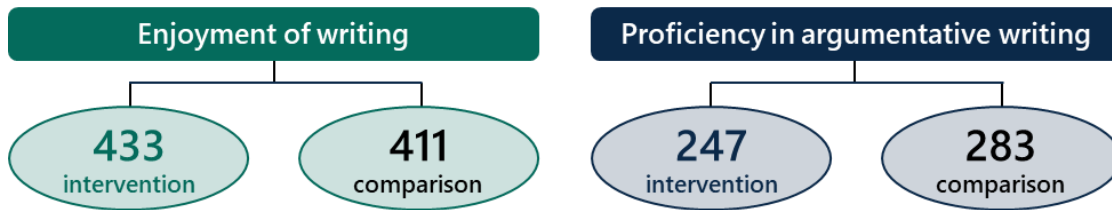
The study team addressed missing baseline values for LWS items and essay scores by imputing with the classroom average score at baseline.

The analytic sample was not balanced between intervention and comparison groups on observable demographic characteristics. For example, across groups, there were differences in the percentage of students who were Black, Latino, English language learners and attending middle school. Overall, about a quarter of students in the final analytic sample were Black or Latino, which were communities in focus for this study.

---

[4] The study team also examined the impact of Ecree on students' confidence in writing and students' proficiency in argumentative writing, as scored by Ecree's algorithmic scoring system.

[5] Essays did not receive a valid score if they were off topic or did not contain enough content to be scored.

Student impact samples

Enjoyment of writing		Proficiency in argumentative writing	
433 intervention	**411** comparison	**247** intervention	**283** comparison

Estimating the impacts of Ecree on student outcomes

The study team used propensity score matching to estimate the impacts of Ecree. A large proportion of randomly assigned teachers and their students did not participate in data collection activities, and those who did participate in the intervention and comparison groups were dissimilar, on average, in terms of measured characteristics. Simply comparing outcomes between the groups that participated in the intervention and comparison groups could produce biased impact estimates if the differences in the characteristics of the two groups (observed or unobserved) also influence the outcomes measured by the study. To address this concern, the study team used propensity score matching which is a way to statistically weight the intervention and comparison groups so that they look similar on characteristics measured at the beginning of the study. This approach increases the likelihood that observed differences in outcomes across the two groups are attributable to having access to Ecree and not differences in their characteristics.

The study team used a propensity score model to estimate the probability of each student being assigned to the intervention group. The model included baseline values of outcomes, an indicator if the baseline value was imputed (one for each outcome), and the student's age, gender, race and ethnicity (Black and/or Latino), and special education status. The study team estimated the propensity score model separately for the sample of students with survey outcomes and the sample of students with essay outcomes.

The team used a weighted least squares regression model to estimate impacts. In the model, intervention students received a weight equal to 1 divided by the estimated probability of being assigned to the intervention group, and comparison students received a weight of 1 divided by the estimated probability of being assigned to the comparison group. In addition to the treatment variable, the model included the baseline value of the outcome and the student's age. The study team clustered standard errors at the teacher level to account for the grouping of students in teachers' classrooms.

Estimating the impacts of Ecree by baseline writing scores[6]

The study team estimated the impact of Ecree by student performance on the writing assessment at the beginning of the study. Students with an overall Smarter Balanced score of 3 or lower were defined as "students with lower scores at baseline," and students with scores of 4 or higher were defined as "students with higher scores at baseline."[7] The study team estimated the impact of Ecree separately for each group

---

[6] The study team also estimated the impact of Ecree by gender and by district.

[7] The analytic sample for survey outcomes among students with lower scores at baseline included 234 students in the intervention group and 251 students in the comparison group; among students with higher scores at baseline, the sample included 199 students in the intervention group and 160 students in the comparison group. The analytic sample for essay outcomes among students with lower scores at baseline included 116 students in the intervention group and 173 students

by estimating a propensity score model for each group and using the resulting weights in the same weighted least squares regression specification as the main model.

Using probability statements to interpret impact estimates on student outcomes

The study team used Bayesian probability statements to interpret the impact estimates from the study. Bayesian analysis is a calculation of the probability that an intervention had a meaningful impact. It estimates the distribution of true impacts given (1) the impacts estimated using study data and (2) the impacts of similar interventions estimated in prior studies. For more information on Bayesian analysis, see [The BASIE (BAyeSian Interpretation of Estimates) framework for interpreting findings from impact evaluations: A practical guide for education researchers](#).

To define prior impacts of similar interventions on student outcomes, the study team referred to impact studies of educational interventions on middle school English language arts achievement that were reviewed by the What Works Clearinghouse and met Group Design standards with or without reservations.[8] Using these prior impacts of similar interventions, along with the Ecree impacts estimated from this study's data, the study team calculated the distribution of true impacts—that is, the likely range of true impacts that could have led to the impacts estimated from the Ecree study, given what we know from prior research. From the distribution of true impacts, the study team then calculated several probabilities representing the likelihood that the impact of Ecree was greater than a specified threshold (for example, they calculated the probability that the impact of Ecree was greater than –0.2, –0.1, 0, 0.1, and 0.2 standard deviations).

For reporting findings in the briefs, we considered a positive impact *likely* if the probability that the impact was greater than zero was 75 percent or above; *potentially likely* if it was between 61 and 74; and unlikely if the probability was 60 percent or less. We chose 75 percent because the intervention we studied is relatively low cost and unlikely to do harm, but we would recommend using a higher confidence level for interpretation (such as 85 or 90 percent) for a higher cost or higher risk intervention.

---

in the comparison group; among students with higher scores at baseline, the sample included 131 students in the intervention group and 110 students in the comparison group.

[8] The procedures and standards handbooks used by the WWC for reviewing studies are available at [WWC | Handbooks and Other Resources (ed.gov)](#).

3 Analytic methods for assessing the implementation of Ecree, how students and teachers used and perceived the tool, and its cost

The study team collected information on student and teacher use of Ecree, student and teacher perceptions of Ecree, the facilitators and barriers to implementation, and the cost of Ecree. The study team used quantitative and qualitative descriptive methods to analyze these implementation data.

Aspects of implementation examined

- Usability
- Usefulness
- Facilitators and barriers
- Cost
- Fidelity

Implementation measures[9]

To construct the implementation measures for analysis, the team used the data collected from student and teacher surveys, teacher interviews, and usage data from the Ecree platform. The survey instruments included the following measures, all of which use a 4-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree):

Students' perception of the usability of Ecree. A study-created item measured students' perception of how easy it was to use Ecree.

Students' perception of the usefulness of Ecree. A study-created scale with two items measured students' perceptions of how much Ecree helped them to learn to write well and improve their writing.

Teachers' perception of the usability of Ecree. A study-created item measured teachers' perception of how easy it was to incorporate Ecree into the classrooms.

Teachers' perception of the usefulness of Ecree. A study-created item measured teachers' perception of how useful Ecree was in helping students improve their writing.

Analyzing other aspects of implementation

To understand Ecree's usability and usefulness in a classroom setting, the study team relied on data from student and teacher surveys data and teacher interviews. The team conducted descriptive analyses of student and teacher survey data by calculating the percentage of respondents who agreed or strongly agreed with the individual items. The student analytic sample sizes for these descriptive analyses varied due to missing item-level data and skip patterns in the survey and ranged from 367 to 368 students in the intervention group. The team used thematic analysis to analyze the responses to teacher interview questions. The team started with descriptive coding using pre-determined codes and refined codes throughout the process. They then conducted an exploratory analysis to identify patterns and develop findings.

To understand facilitators and barriers to the tool's implementation, the study team relied on teacher interviews. The team applied qualitative analytic approaches to identify implementation facilitators and barriers to implementation, including those related to COVID-19.

---

[9] The study team also measured Ecree's suitability for students from diverse backgrounds, students' perceptions of classroom climate, teachers' attitudes on using technology, and frequency of writing tasks assigned by teachers.

The study team also estimated the cost of delivering Ecree in this study. The cost calculation accounts for Ecree's software license fee, the compensation teachers received for attending the hour-long webinar training offered by Ecree support staff, and "soft" costs associated with using Ecree: (1) time for teachers to use the tool, (2) time for district network administrators to integrate and install the tool, and (3) time for district network administrators to communicate with teachers about the tool. The cost of time for teachers, district leaders, and district network administrators are not fees paid to Ecree. Rather, they are estimates of the value of their time spent implementing Ecree. Cost calculations assume students have access to a computer or tablet and high-speed internet in schools and do not account for any related expenditures districts might incur to provide those supports. Ecree also provides limited support to teachers and districts, which is included as part of the subscription fee. The total cost to use Ecree was converted to a cost per student based on the 958 students who were offered Ecree.

Assessing fidelity of implementation

The study team examined the degree to which teachers and students with access to Ecree (i.e., those in the intervention group) maintained fidelity to the criteria that the developer identified as essential to implementation (Table 2). To assess fidelity, the study team conducted descriptive analyses of usage data from Ecree's platform and teacher attendance records that contained information about when users completed activities in Ecree and which teachers attended training. The team calculated the percentage of teachers and students who met each criterion, as well as the percentage who met all the criteria.

Table 2. Ecree implementation fidelity criteria

Activity	Teachers	Students
Essay assignment and completion activities		
Essays	Assigned six essays	Completed six essays (with at least one revision)
Teacher training		
Training session	Attended webinar training	n.a.

Interested in learning more about the Ecree study? Email tmusaddiq@mathematica-mpr.com.

Interested in implementing Ecree in the classroom? Email contact@ecree.com.

Let's Progress Together. Contact…

mathematica.org