

Study Methods for Briefs about MI Write Research Findings

This summary provides details about a study of MI Write, an automated writing feedback tool, which took place during the 2021–2022 school year. The MI Write team and an evaluator from the University of Delaware (Joshua Wilson, Ph.D.), with technical assistance from Mathematica, designed and conducted the study to examine teachers’ and students’ use of the tool during the study and the impact of MI Write on student and teacher outcomes. This summary accompanies a series of briefs summarizing results of the MI Write 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](#).

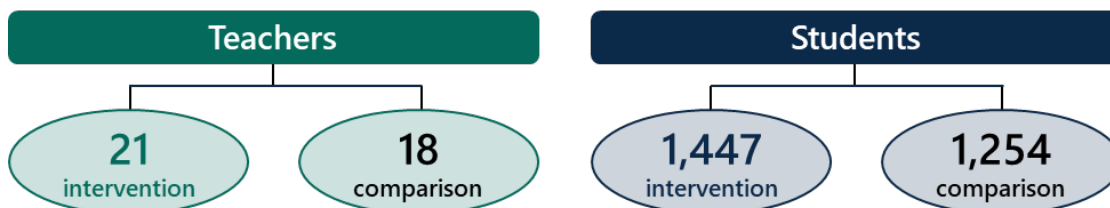
1 Study design

The study team designed and carried out a randomized controlled trial to examine the impact of MI Write on student and teacher outcomes. The team also collected information to understand the usability, usefulness, and utilization of MI Write, its suitability for students from diverse backgrounds, and the facilitators and barriers to implementing the tool.

Sample and random assignment

The study team recruited three school districts in New Jersey and North Carolina, including one rural, one urban, and one suburban school district, to participate in the study. Teachers who taught grades 7 and 8 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 MI Write (intervention group) or to continue typical instructional practices (comparison group). Across districts, 21 teachers were assigned to the intervention group and 18 teachers were assigned to the comparison group. There were 1,447 students taught by teachers in the intervention group and 1,254 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 1,369 students taught by teachers in the intervention group and 1,247 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 MI Write study, email joshwils@udel.edu.

Data collection

Data sources for the MI Write study included study-administered student and teacher surveys, study-administered student writing assessments, and usage data collected from the MI Write platform. Table 1 summarizes the types of data obtained from these sources and the response rates for each.²

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	Enjoyment of and confidence in writing; usability, usefulness, and cultural responsiveness of MI Write	2,364 (88%)	1,196 (83%)	1,168 (93%)
Follow-up student writing assessment	Proficiency in argumentative writing	2,208 (82%)	1,089 (75%)	1,119 (89%)
Follow-up teacher survey	Confidence in teaching writing; use of evidence-based practices for teaching writing; usability and usefulness of MI Write, and its suitability for students from diverse backgrounds; usability and usefulness of professional development supports; facilitators and barriers to implementation	37 (95%)	19 (90%)	18 (100%)
Teacher interviews and focus groups	Facilitators and barriers to implementation	9 (n.a.) ^a	9 (n.a.) ^a	n.a. ^b
Usage data collected from the MI Write platform, teacher attendance records and coaching logs	Student and teacher implementation fidelity	1,260 students (87%), 19 teachers (90%)	1,260 students (87%), 19 teachers (90%)	n.a. ^c

Note: Thirty-nine teachers were randomly assigned to a group (21 intervention-group and 18 comparison-group teachers). A total of 2,701 students (1,447 intervention-group and 1,254 comparison-group students) were enrolled in the classrooms of teachers who were randomly assigned. These numbers serve as denominators in response rate calculations.

^a 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. Interviews were conducted both individually and in small groups.

^b The study team did not interview teachers in the comparison group because they did not implement MI Write.

^c Students and teachers in the comparison group did not have MI Write accounts and did not use MI Write.

n.a. = not applicable.

² The study team also administered a beginning-of-year (baseline) survey to students and teachers, administered a baseline writing assessment to students, collected administrative data on student demographics from participating school districts, and conducted classroom observations of teachers. We do not describe those data sources here because findings based on the samples from those data sources are not included in the briefs.

2 Analytic methods for estimating impacts of MI Write on student outcomes

The study team examined the impact of MI Write on several student outcomes, including students' enjoyment of writing, confidence in writing, and proficiency in argumentative writing. The team used regression analysis to estimate the impacts of MI Write on these outcomes.

Student outcome measures³

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 (0 = strongly disagree, 1 = disagree, 2 = agree, 3 = strongly agree) across four items. The study estimated the impact of MI Write on an overall LWS score that was constructed by taking the average value across the four items.

Students' confidence in writing. The Self-Efficacy for Writing Scale (SEWS) measured students' confidence in writing across three subscales (Ideation, Conventions, and Self-regulation) on a scale of 0 (low confidence) to 100 (high confidence) across 19 items. The study team estimated the impact of MI Write on an overall score for each subscale that was constructed by taking the average value of the items in that subscale.

Students' proficiency in argumentative writing. Writing assessments in the form of essays measured students' proficiency in argumentative writing. Human raters scored the 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 MI Write 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: [Menu of Measures: Secondary Writing \(mathematica.org\)](https://mathematica.org).

Teacher outcome measures

To construct the teacher outcome measures for the study, the team collected data through teacher surveys. The teacher survey instrument incorporated the following measures:

Teachers' confidence in teaching writing. The National Survey of Teachers' Preparation and Practices in Teaching Writing (TPPTW) – Teacher Efficacy for Writing subscale measured teachers' confidence in teaching writing on a 4-point Likert scale (0 = strongly disagree, 1 = disagree, 2 = agree, 3 = strongly agree) across nine items. The study team estimated the impact of MI Write on an overall score that was constructed by taking the average value across the nine items.

³ The study team also examined the impact of MI Write on students' beliefs that writing is a recursive process and students' proficiency in argumentative writing, as scored by MI Write's Project Essay Grade automated scoring system.

Teachers' use of evidence-based practices for teaching writing. The TPPTW – Teacher Use of Evidence-Based Writing Practices subscale measured the frequency with which teachers used various teaching practices. The scale ranged from 0 (never) to 7 (several times a day) and was used across 15 items. The study team estimated the impact of MI Write on an overall score that was constructed by taking the average value of the 15 items.



For more information on the teacher measures used in this study, see the Menu of Measures [[Menu of Measures: Secondary Writing \(mathematica.org\)](https://www.mathematica.org)].

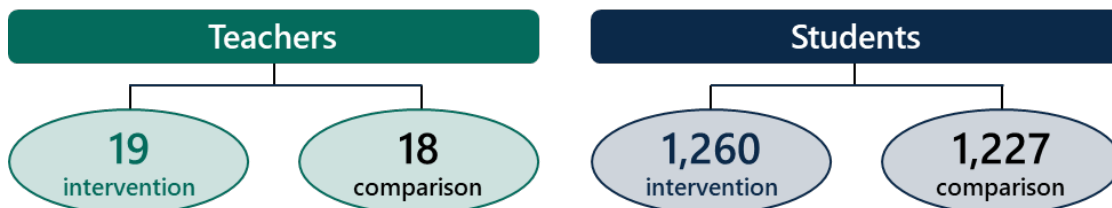
Analytic samples

The final analytic sample for all student outcomes included students whose parent consented, whose teacher remained in the study, and who had observed outcome data for at least one outcome. The sample included 1,260 students in the intervention group and 1,227 students in the comparison group, for a total of 2,487 students. Two of the 39 randomly assigned teachers dropped out of the study before participating in the beginning-of-year (baseline) data collection. For the remaining 37 teachers and their students, the study team used a multiple imputation process to address missing values when baseline and follow-up data were not complete.⁴

Intervention and comparison groups in the student analysis sample were balanced on the observable demographic characteristics of gender and English learner status, but they differed in the percentage of students who were Black, Latino and/or experiencing poverty (the communities in focus for the study). They also differed by the percentage who were receiving special education services. Across all districts, 80 percent of students in the study were members of at least one the communities in focus. Specifically, 17 percent of students were Black, 45 percent of students were Latino, and 44 percent received free or reduced-price lunch.

The final analytic sample for all teacher outcomes included 19 teachers in the intervention group and 18 teachers in the comparison group. There were no missing data for the analyses of teacher outcomes. The teacher sample used in the analysis was balanced between intervention and comparison groups on all observable characteristics for which data was collected, including gender, race and ethnicity, teaching experience, and education.

Analytic samples



⁴ The study team used SPSS V.28's fully conditional Markov chain Monte Carlo algorithm to complete multiple imputation. The study team imputed Smarter Balanced trait scores for essays that did not receive a valid score (that is, they were off topic or did not contain enough content to be scored) and overall survey scale scores that were left blank.

Estimating the impacts of MI Write on student and teacher outcomes⁵

To estimate the impact of MI Write on student outcomes, the study team used an ordinary least squares regression model in which students were coded as belonging to the intervention and comparison groups based on their originally assigned conditions. In addition to the treatment variable, the model included fixed effects for random assignment blocks, the baseline measure of the outcome, whether the student was a member of one of the communities in focus, and whether the student received special education services. The study team clustered standard errors at the teacher level to account for the grouping of students in teachers' classrooms.

To estimate the impact of MI Write on teacher outcomes, the study team used an ordinary least squares regression model in which teachers were coded as belonging to the intervention and comparison groups based on their originally assigned conditions. In addition to the treatment variable, the model included fixed effects for random assignment blocks and the baseline measure of the outcome. Teacher demographic characteristics were balanced across groups, so the study team did not include them as covariates in the model.

Using probability statements to interpret impact estimates on student and teacher 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.⁶ To define prior impacts on teacher outcomes, the study team referred to impact studies of educational interventions on miscellaneous outcomes, including teacher outcomes, that were reviewed by the What Works Clearinghouse and met Group Design standards with or without reservations. Using these prior impacts of similar interventions, along with the MI Write 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 MI Write 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 MI Write was greater than a specified threshold (for example, they calculated the probability that the impact of MI Write 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

⁵ The study team also estimated the impact of MI Write by student membership in the communities in focus and by district.

⁶ The procedures and standards handbooks used by the WWC for reviewing studies are available at [WWC | Handbooks and Other Resources \(ed.gov\)](#).

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.

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

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

Implementation measures⁷

To construct the implementation measures for analysis, the team used the data collected from student and teacher surveys, individual and small group teacher interviews, and usage data from the MI Write platform. The survey instruments included the following measures, all of which use a 4-point Likert scale (0 = strongly disagree, 1 = disagree, 2 = agree, 3 = strongly agree) unless otherwise noted:

Students' perception of the usability of MI Write. A study-created scale with six items measured students' perceptions of how easy it was to use specific features of MI Write and the system overall.

Students' perception of the usefulness of MI Write. A study-created scale with six items measured students' perceptions of how much MI Write helped them with specific aspects of the writing process.

Students' perception regarding the degree to which MI Write is suitable for students from diverse backgrounds. A study-created scale measured students' perceptions of MI Write's cultural responsiveness, or the degree to which it is suitable for students with diverse abilities and identities. The 4-point Likert scale (0 = strongly disagree, 1 = disagree, 2 = agree, 3 = strongly agree) was used across four items.

Teachers' perception of the usability of MI Write. A study-created scale with seven items measured teachers' perceptions of how easy it was to use specific features of MI Write and the system overall.

Teachers' perception of the usefulness of MI Write. A study-created scale with 11 items measured teachers' perceptions of how much MI Write helped them with specific aspects of teaching writing and helped students.

Teachers' perception of the usability of the professional development supports. A study-created scale with five items measured teachers' perceptions of how easy it was to engage with and use specific aspects of the professional development.

Aspects of implementation examined

- Usability
- Usefulness
- Suitability for diverse students
- Facilitators and barriers
- Cost
- Fidelity

⁷ The study team also measured students' and teachers' perceptions of the social desirability of MI Write.

Teachers' perception of the usefulness of the professional development supports. A study-created scale with six items measured teachers' perceptions of the usefulness of specific aspects of the professional development.

Teachers' perception of the barriers to implementing MI Write. A study-created scale with 11 items measured teachers' perceptions of the extent to which specific factors were barriers to implementing MI Write in their classroom. A scale of 0 (not a barrier) to 3 (great barrier) was used.

Analyzing other aspects of implementation

To understand MI Write's usability and usefulness in a classroom setting and the degree to which the tool is suitable for students from diverse backgrounds, the study team relied on data from student and teacher surveys, and individual and small group 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 in the survey scales. They also calculated the percentage of respondents whose average value across the items in each scale was a 2 or higher (where 2 corresponded to agree and 3 corresponded to strongly agree). The student analytic sample sizes for these descriptive analyses varied due to missing item-level data and ranged from 1,182 to 1,187 students in the intervention group. The team also used qualitative analytic approaches to examine teacher responses to interview questions to identify patterns and central themes.

To understand facilitators and barriers to the tool's implementation, the study team relied on teacher interviews and surveys, as well as coaching logs. The team applied qualitative analytic approaches to identify implementation facilitators and barriers to implementation, including those related to COVID-19. To further understand implementation barriers, the team also used descriptive methods to analyze coaching logs and teacher survey data about the extent to which various factors were barriers. The analytic sample for teacher surveys and coaching logs included 19 teachers in the intervention group.

The study team also estimated the cost of delivering MI Write in this study, based on its software licensing fee and the time for MI Write staff to provide training and coaching for study teachers.

Assessing fidelity of implementation

The study team examined the degree to which teachers and students with access to MI Write (i.e., those in the intervention group) maintained fidelity to the core training, coaching, and usage criteria that the developer identified as essential to implementation (Table 2). To assess fidelity, the study team conducted descriptive analyses of usage data from MI Write's platform, coaching logs, and teacher attendance records that contained information about when users completed activities in MI Write and when teachers attended training and professional development sessions. 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. MI Write implementation fidelity criteria

Activity	Teachers	Students
Essay assignment and completion activities		
Prewriting activities	Assigned eight prewriting activities to students	Completed eight prewriting activities
Essays	Assigned eight essays to students	Completed eight essays (with at least two revisions)
Interactive lessons	Assigned eight interactive lessons to students	Completed eight interactive lessons
Peer reviews	Assigned three peer reviews to students	Completed three peer reviews
Annotated feedback	Completed annotated feedback for 100% of essays after January	n.a.
Teacher training and professional development		
Training session	Attended initial training session	n.a.
Professional learning sessions	Attended three professional learning sessions	n.a.
Coaching sessions	Attended five coaching sessions	n.a.

The MI Write team (Corey Palermo, Ph.D., Halley Eacker, Ph.D., and Jessica Coles) and University of Delaware evaluator (Joshua Wilson, Ph.D.) designed and conducted the study with technical assistance from Mathematica ([Ryan Ruggiero](#), [Lindsay Fox](#), and [Megan Shoji](#)). Mathematica ([Lindsay Fox](#) and [Marykate Zukiewicz](#)) wrote the brief with contributions from the MI Write and UD teams. [Virginia Knechtel](#) reviewed the content and provided feedback. This publication was prepared for the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

Interested in learning more about the MI Write study? Email joshwils@udel.edu.

Interested in implementing MI Write in the classroom? Email info@miwrite.net.

Let's Progress Together. Contact...