Use of Formative by Newsela Fast Cycle Feedback to Support Student Math Outcomes

Prepared for:
Formative by Newsela

Prepared by LearnPlatform by Instructure:
Ashley Hunt, Ph.D., Senior Researcher
Chelsae Long, M.A., Junior Researcher

January 2024

EXECUTIVE SUMMARY

The aim of this study was to examine the relationship between fast cycle feedback and student math performance. Formative by Newsela contracted with LearnPlatform by Instructure, a third-party edtech research company, as part of the Futures Forum Learning Engineering Tools Competition (Phase II – Assessment).

Study Sample and Measures

This study occurred during spring 2023 and included 1,524 students from grades 5 through 8 from multiple school districts. Researchers used results from state standards-aligned Formative math assessment scores as the student achievement outcome. Analyses included descriptive statistics and partial correlations at the student and item-levels. Researchers examined whether prompting educators to give feedback while completing a Formative assessment was related to better scores on the post-test.

Findings

On average, students in the feedback group performed better on the follow up assessment. Specifically, feedback group assignment was associated with a significant, moderate effect (Hedge's g = 0.33) on student post-test performance. This indicates a student in the comparison group that performed at the 50th percentile would have been expected perform at the 63rd percentile in the feedback group. Subsequent grade-level analyses found larger effect sizes among grade 5 (g = 0.56) and grade 7 (g = 0.71) students.

Key Findings
Students in the fast cycle feedback group outperformed peers who did not receive feedback.

+ Impact on Average Student Performance: An average student who did not receive fast cycle feedback would be expected to score 13 percentile points higher on the outcome assessment had they received feedback.

+ Grade-level differences: Analyses found larger effect sizes among students in grades 5 and 7.

Note: These findings were statistically significant at the p = 0.05 level.

Conclusions

Results support the project goal of showcasing how frequent, multimodal, and rapid feedback in the Formative platform can support student learning outcomes. Furthermore, the study provides results to satisfy ESSA evidence requirements for Level III (*Promising Evidence*) given the study design and positive, statistically significant findings.

TABLE OF CONTENTS

FIGURES

Introduction

Formative contracted with LearnPlatform by Instructure, a third-party edtech research company, to examine the relationship between Fast cycle feedback and student math performance as part of the Futures Forum Learning Engineering Tools Competition (Phase II – Assessment).

Research has shown timely, constructive feedback promotes positive learning growth in the classroom (Hattie, 2008). Furthermore, research suggests that as education technology advances, it is critical to understand whether and how tools for giving feedback relate to student performance (Hirsh-Pasek et al., 2015). Formative allows teachers to view and act on student work in real time. With fast cycle feedback, teachers can leave any modality of feedback, including recorded audio or video, drawing directly on top of student work, or text.

The present study had the following research questions:

1. Did students who received feedback on their pre-test perform better on the follow up assessment?
2. Did students who exchanged more feedback messages with their teacher perform better on the post-test?

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

Study Procedures and Timeline

This study occurred during spring 2023 school year. All middle school math educators that were already users of the Formative platform were invited to participate in the study. Those who joined were randomly assigned to the feedback or no feedback group and asked to administer a pre- and post-test to their students on the Formative platform. Participants administered the post-test within two weeks of completing the pre-test.

The pre- and post-assessments were comprised of items that aligned with state math standards for students in grades 5 through 8. Both the pre- and post-assessments included the same number of questions and assessed the same content within each classroom. Analyses were conducted at the assessment and item-level (see Appendix A).

Analyses were conducted using IMPACT (Integrated Metrics Producing Analytics on Classroom Technologies), the LearnPlatform by Instructure rapid cycle evaluation engine. IMPACT conducts analysis based on the sampling technique embedded in the research design. The platform enables a full range of sampling methods and techniques to analyze a sample of students or schools representative of the population of interest. Further, IMPACT employs methods (e.g., statistical adjustments and inclusion of covariates) that maximize representativeness.

Program Implementation

The results below summarize Formative usage during spring 2023 based on internal usage data. Only students who completed both a pre- and post-assessment were included for analyses.

Student-Level Usage & Feedback

The results below summarize overall Formative platform usage and feedback group assignment from the study. Of the 1,524 students included in the analyses, nearly one third (32%) were assigned to the feedback condition[1] (*n* = 484, see Figure 1).
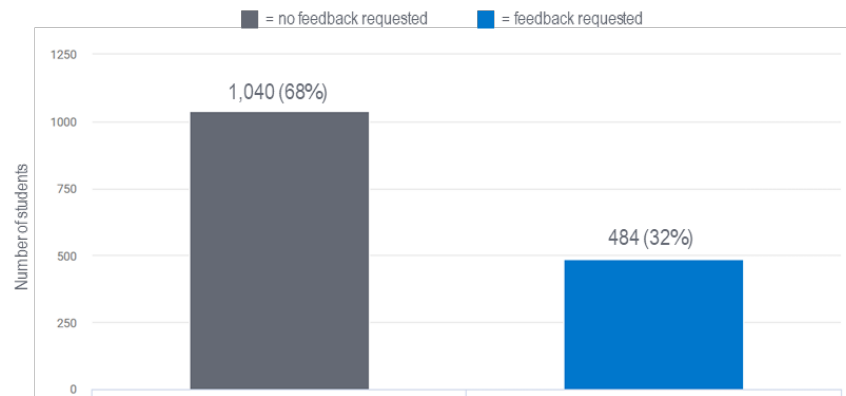

Figure 1. Number of students by feedback group assignment

Figure 2 below summarizes grade-level usage. The study included 241 grade 5 students, 10% of which were assigned to the feedback condition. Grade 6 analyses included 328 students with 25% in the feedback condition. Grade 7 had the most balanced assignment with 56% of the 339 students appearing in the feedback condition. Finally, 616 grade 8 students were in the study, 31% of which were in the feedback condition.
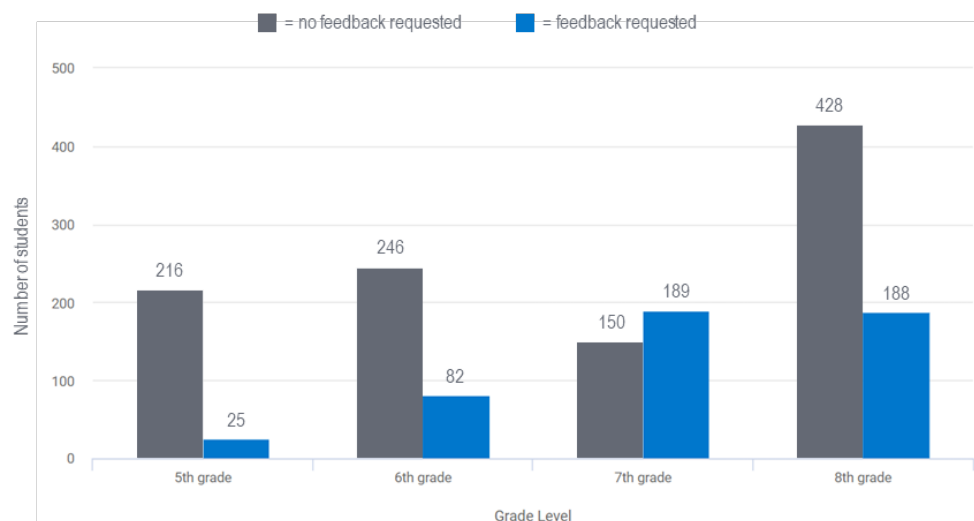

Figure 2. Number of students by feedback group assignment and grade-level

---

[1] An equal number of invitations to participate in this study were sent to math teachers registered on Formative. Invitations were not segmented by grade level and not all teachers opted to participate in the study. Together, this resulted in uneven study participation across conditions and grades.

Figure 3 below reports how many feedback messages were exchanged on the Formative platform between students and educators in the feedback condition (*n* = 484). Off-platform feedback communication may have taken place, but was not captured in these analyses.
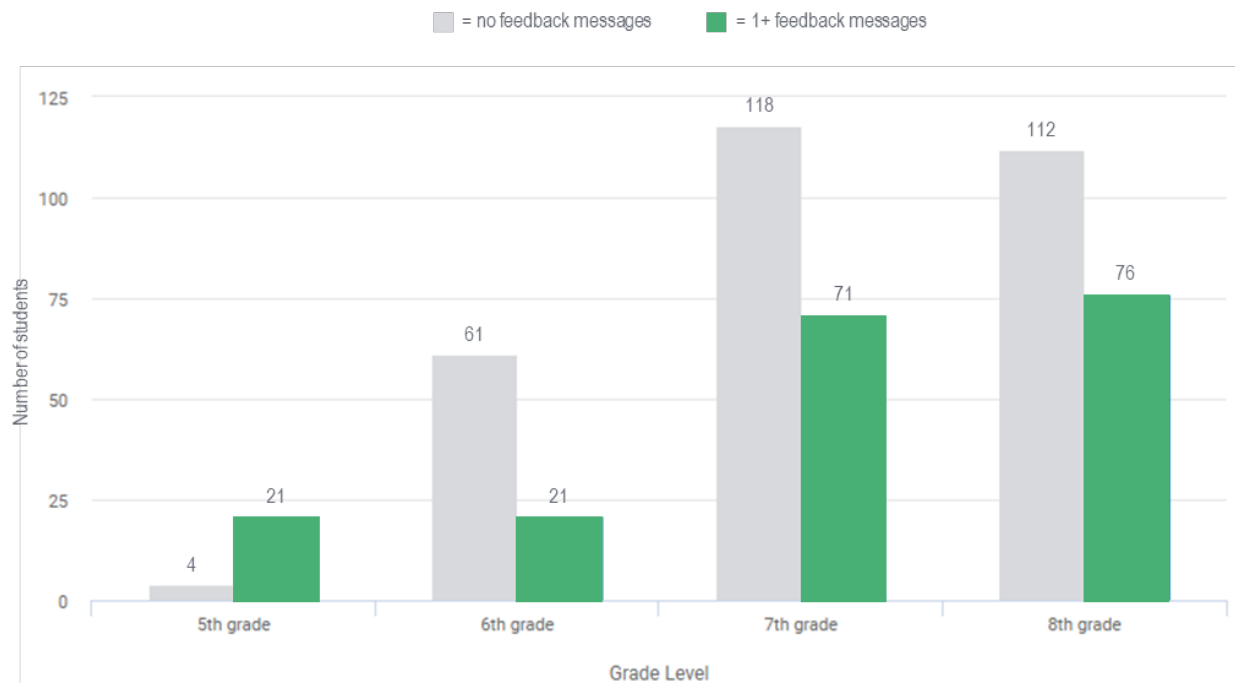

Figure 3. Number of on-platform messages exchanged by grade-level

The majority (84%) of the 25 grade 5 students in the feedback condition exchanged at least one on-platform message. One in four of the 82 grade 6 students (25%) exchanged one or more messages on the platform. Among the 189 grade 7 students in the feedback condition, 38% exchanged at least one message. Nearly half (40%) of the 188 grade 8 students in the feedback condition engaged in communication on the platform. The low level of feedback exchange on platform in grades 6, 7, and 8 are important to consider when interpreting study results.

Findings

The results below summarize results from analyses comparing students in the two feedback conditions. Only students who completed both a pre- and post-assessment (n = 1,524) were included for analyses. Researchers report statistically significant findings at the $p < .05$ level. Statistically significant findings are green to indicate a positive correlation or red to indicate a negative correlation. Findings that are not statistically significant are shown in yellow.

Researchers used Hedge's g (Hedges, 1981) to characterize the practical importance of statistically significant effects of Formative, which typically range from -2 through +2. A Hedge's g value of 0.15 indicates a small effect, while a value of 0.40 indicates a moderate effect. A value of 0.75 or above is considered a large effect. Hedge's g values were also converted using the What Works Clearinghouse Improvement Index (WWC, 2022) to report effects in terms of percentile improvement between groups.

Student-Level Analysis

Researchers conducted partial correlations examining the association between feedback group and student post-test scores controlling for pre-test performance. Results indicated a positive, statistically significant relationship ($p < .001$), such that students who were assigned to receive feedback on their pre-assessment had higher post-assessment performance (see Figure 4).

The Hedge's g value and 95% confidence interval indicate that the result represents a statistically significant small to moderate effect (g = 0.33). In other words, if a student in the comparison performed at the 50th percentile on the post-test, they would have been expected to score in the 63rd percentile in the feedback group.
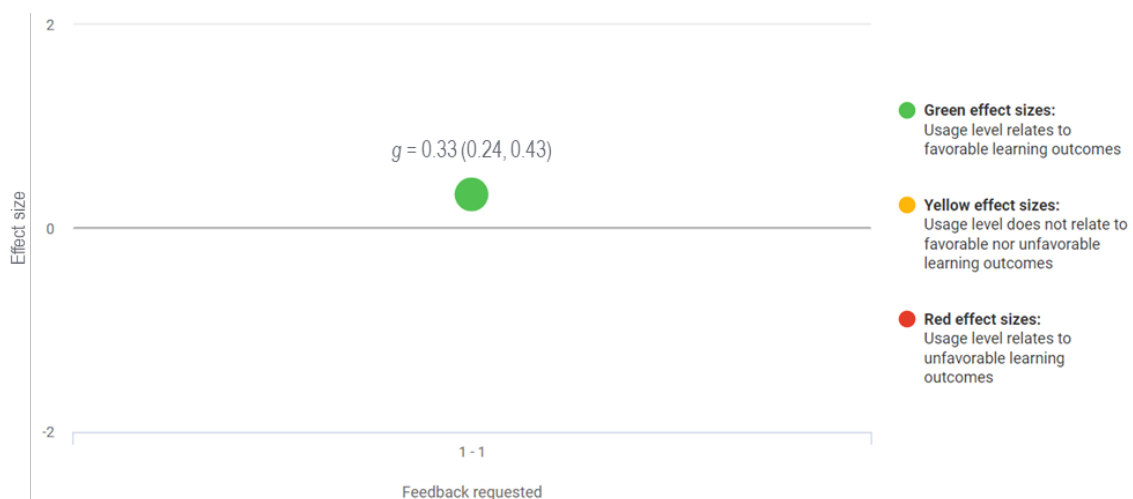

Figure 4. Student-level relationship between feedback group assignment and performance

Subgroup analysis by grade level revealed a consistent pattern of positive results among students in grades 5, 7, and 8 (see Figure 5). The Hedge's g values and 95% confidence intervals indicate a moderate effect in grade 5 (g = 0.56, a 21 percentile point improvement), a large effect in grade 7 (g = 0.71, a 26 percentile point improvement), and a small effect in grade 8 (g = 0.27, an 11 percentile point improvement). Group assignment was not significantly associated with student performance among students in grade 6, possibly due to lower usage of on-platform feedback messages (25%) within that group.
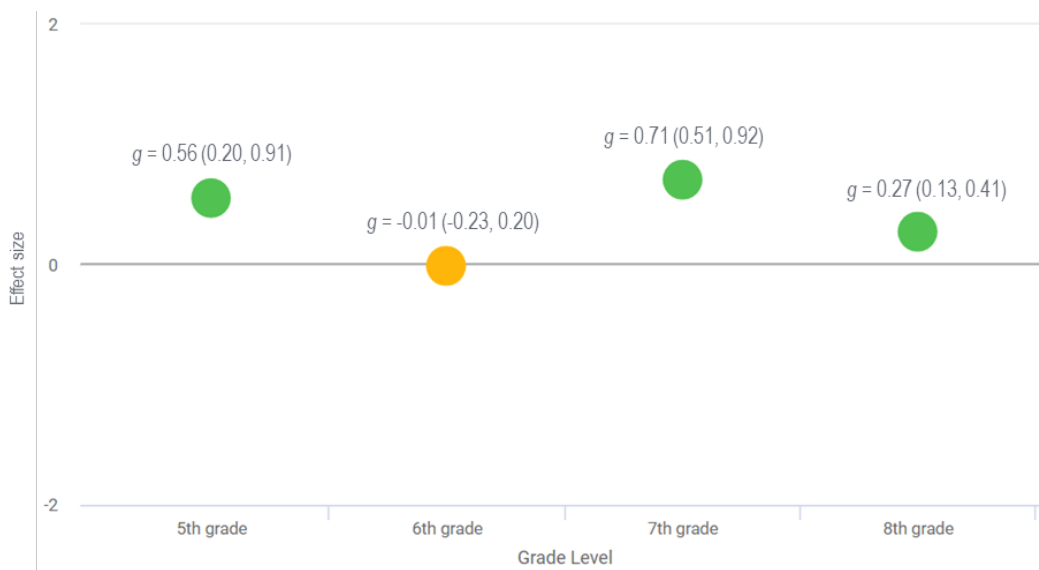

Figure 5. Student-level relationship between feedback group assignment and performance by grade level

Additional partial correlations among the feedback group examined the association between number of feedback messages exchanged and student post-test scores controlling for pre-test performance (see Figure 6). The grade 5 sample size was not sufficient for inclusion of these analyses.

The association between number of messages exchanged and post-test performance was positive and significant indicating a large effect for grade 6 students (g = 0.82, a 29 percentile point improvement). However, the relationship was significantly negative among grade 8 students. The relationship was not significant among grade 7 students. The mixed results suggest that the value of additional feedback messages between students and educators varies by student grade level.
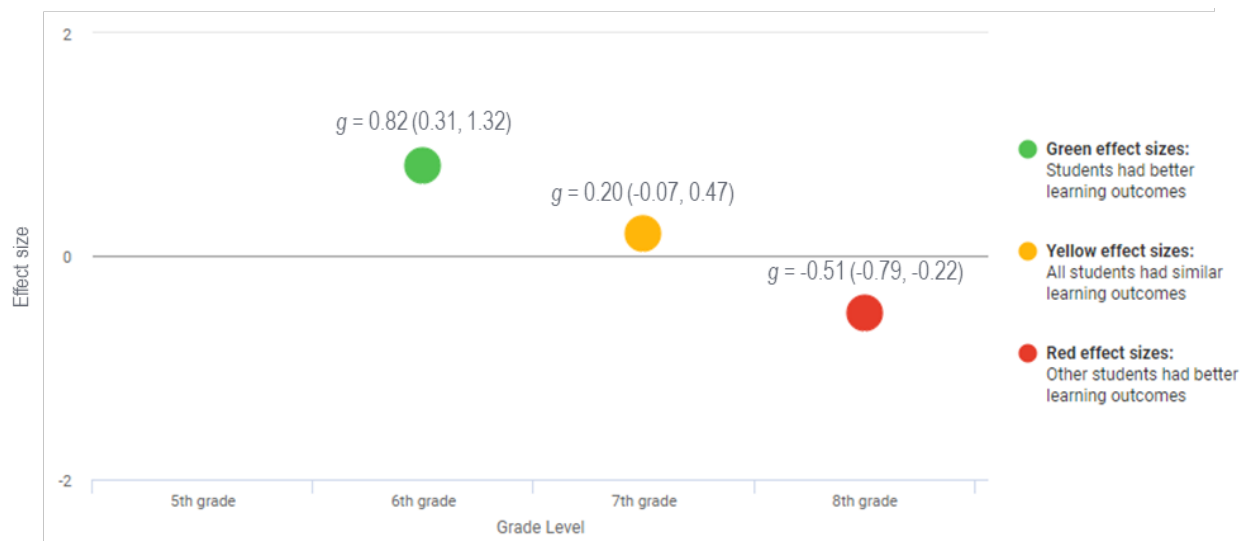
Figure 6. Student-level relationship between number of on-platform messages exchanged and performance

Item-Level Analysis

Researchers conducted partial correlations examining the association between feedback group and correctness of post-test items controlling for pre-test performance. There was a positive, statistically significant relationship ($p < .001$), such that items completed in the feedback condition were more likely to be correct at post-test (see Figure 7). The Hedge's g value ($g = 0.23$) and 95% confidence interval indicate that the result represents a statistically significant small effect.

Figure 7. Relationship between feedback group assignment and post-test item correctness

Item-level analyses of the relationship between feedback group and performance mirrored the student-level results and revealed a consistent pattern of positive results among students in grades 5, 7, and 8 (see Figure 8). The Hedge's g values and 95% confidence intervals indicate a moderate effect in grades 5 (g = 0.42) and 7 (g = 0.46), and a small effect in grade 8 (g = 0.20).
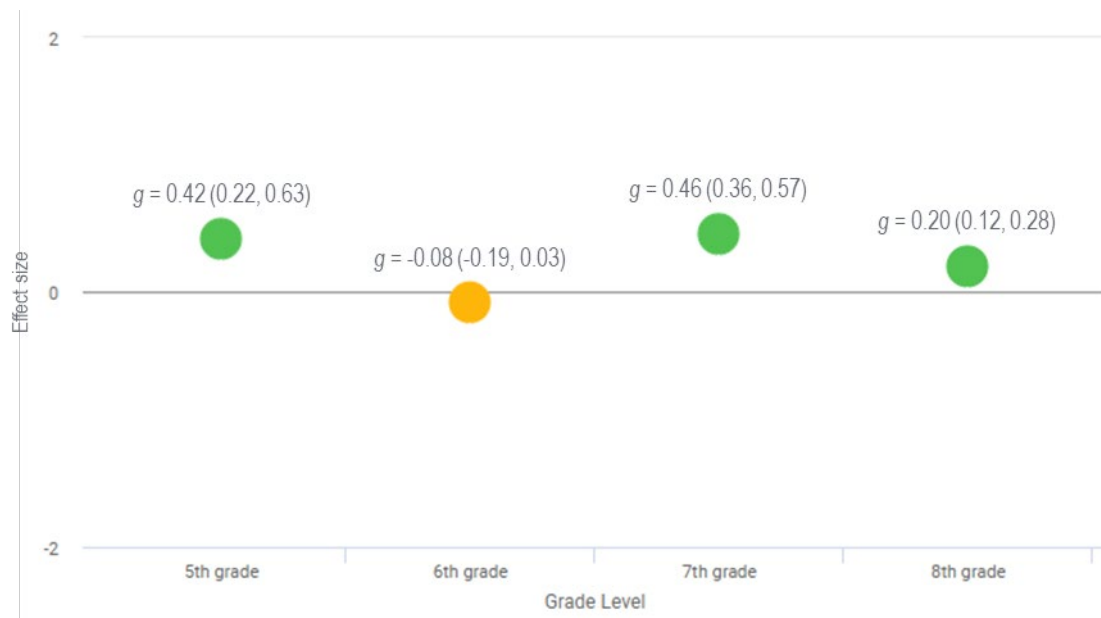

Figure 8. Relationship between feedback group assignment and post-test item correctness by grade level

Recommended Next Steps

For next steps, we recommend that Formative continue to explore the data surrounding Formative feedback implementation and conduct additional research. For example:

- Expanding sampling to other subject areas, grade-levels, and more specific assessment types would help to generate more and varied evidence of the effect of Formative on student performance.

- Results show that many students in the feedback condition did not receive feedback on-platform. Expanding capacity for capturing feedback delivered through different channels would allow for greater understanding of both implementation and the results related to the extent of feedback communication exchange between teachers and students.

- Similarly, data in the current study did not specify whether initial student responses were correct, making it impossible to distinguish between students not receiving and not needing feedback. Expanding this capacity would improve the precision of results.

- A large number of records were excluded from analyses due to missing pre- or post-test data. Integration of more contextual information or other implementation data would allow for greater exploration and understanding of attrition.

Conclusions

Results support the project goal of showcasing how frequent, multimodal, and rapid feedback in the Formative platform improved student learning outcomes. Additional investigation is needed to better specify grade-level differences and explore how the effect of Formative feedback varies by modality.

Furthermore, given positive outcome findings, this study provides results to satisfy Every Student Succeeds Act (ESSA, 2015; USDE, 2016) evidence requirements for Level III (*Promising Evidence*). Specifically, this study met the following criteria for Level III:

- Correlational study

- Proper design and implementation

- Statistical controls through covariates

- At least one statistically significant, positive correlation with statistical controls for selection bias

References

Every Student Succeeds Act, Pub. L. No. 114-95 (2015).
    https://www.govinfo.gov/app/details/PLAW-114publ95.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*.
    Routledge.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related
    estimators. *Journal of Educational Statistics, 6*(2), 107-128.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting
    education in "educational" apps: Lessons from the science of learning. *Psychological
    Science in the Public Interest*, 16(1), 3-34.

United States Department of Education (2016). Non-Regulatory Guidance: Using Evidence to
Strengthen Education Investments.
    https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf.

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).

Appendix A

Invitations to participate in this study were sent to all math teachers registered on Formative. The study sample was comprised of educators and their students in the 27 states highlighted in blue in the map below.
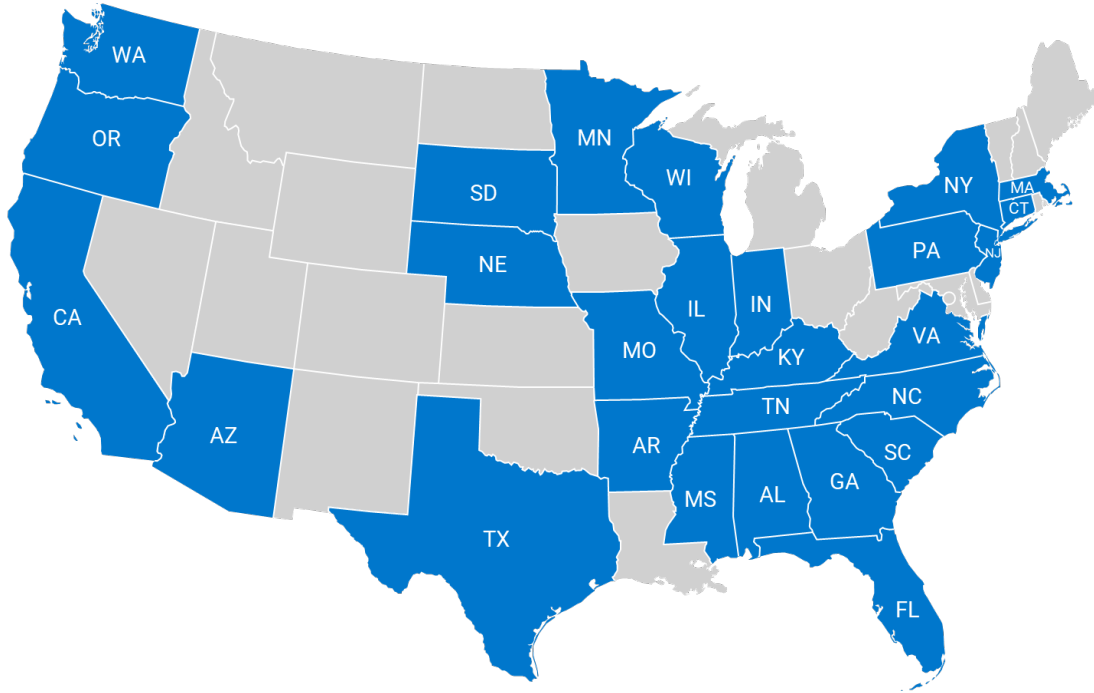

Figure A1. US map indicating states with study participants

The figures below summarize item-level results from the study. Overall, students in grades 5 through 8 completed 7,010 assessment items across the pre- and post-assessments. Results mirrored the student-level results, with one in three items (33%) representing responses from students in the feedback condition.
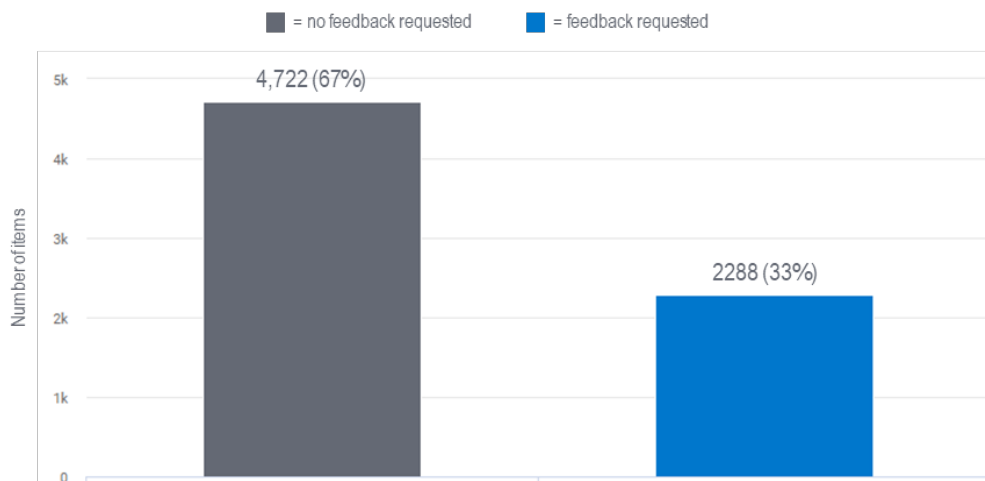

Figure A2. Number of assessment items completed by feedback group assignment

The chart below displays the number of items completed by grade. Grade 5 students answered a total of 925 items, 11% of which were in the feedback condition. Grade 6 analyses included 1,515 items and 25% were in the feedback condition. Grade 7 had the most balanced assignment among the grades with 56% of the 1,604 items appearing in the feedback condition. Finally, grade 8 students responded to 2,966 items, 30% of which were in the feedback condition.
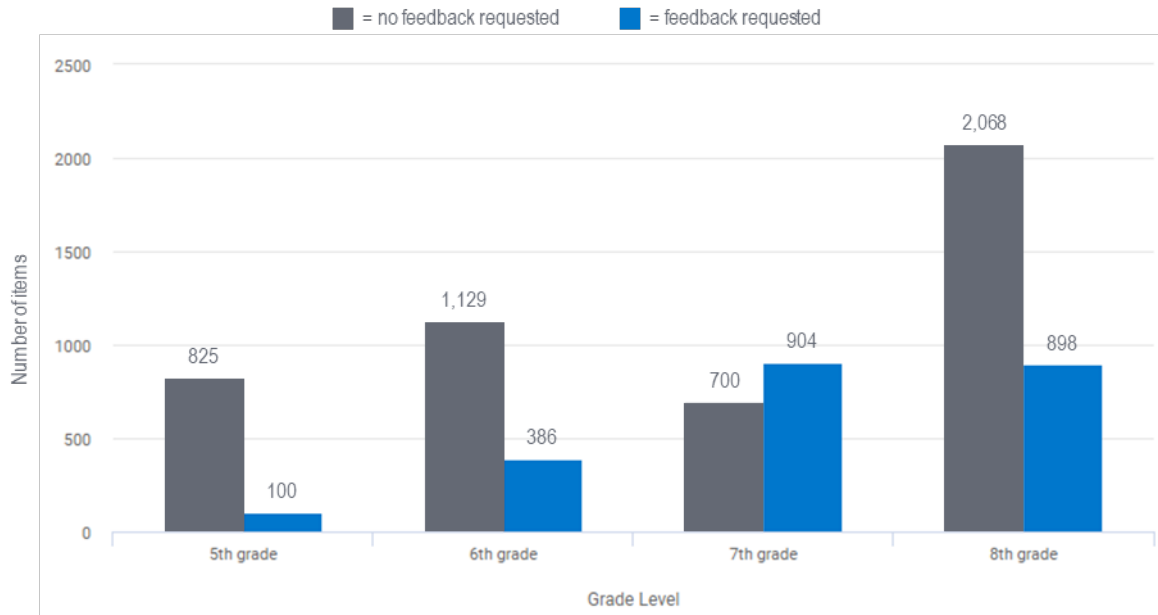

Figure A3. Number of assessment items completed by feedback group assignment and grade level