

Appleton Area School District × Revision History

RESEARCH PILOT — SPRING 2026 — APPLETON AREA SCHOOL DISTRICT

349 Appleton students and 13 teachers were surveyed before and after using Byline (students) and Revision History (teachers) for 8 weeks

KEY TAKEAWAYS

Teachers: Increased trust in student writing authenticity

Students: Decreased concern about being falsely accused of using AI

Students: Increased engagement with writing

Alongside an amazing team in Appleton Area School District, a public school district in Wisconsin, and Dr. Joshua Wilson, Associate Professor in the School of Education at the University of Delaware, we ran a research pilot with 349 students and 13 teachers, gauging teacher and student perception changes before and after implementing Revision History with teachers and Byline with students.

The results were very encouraging when comparing pre-pilot to post-pilot perception changes. Teachers reported a large and statistically significant increase in trust that students were authentically doing their work. Students reported a smaller but statistically significant decrease in concern about being falsely accused of using AI, along with small but statistically significant gains across various measures of engagement with writing.

WHY WE DID THIS RESEARCH

It has never been more difficult for teachers to assess the authenticity of student writing — and, in turn, to gauge a student's level of effort and learning. To help keep students honest and better understand their students' work journey, 200,000+ teachers now use Revision History to make their students' writing process visible.

We built Byline for two reasons:

1. Provide teachers with even more insight into students' writing journey, in addition to what happens inside a Google Doc or Slide (student website visits, Chrome extension activity, AI use)
2. Provide teachers with the ability to put teacher-enabled/disabled, guardrailed, transparent AI in front of students to incentivize and teach responsible AI use

But before rolling out Byline, we wanted to conduct research to ensure we were helping boost trust in classrooms — on both the teacher and student sides.

FINDING 1

Teachers: increased trust in student writing authenticity

The teacher sample is small (13 matched responses), so read these as exploratory, but all three shifts below survived statistical correction. Scores are means on a 1–5 scale.

Survey item	Pre	Post	Change
I have adequate tools to verify writing authenticity	2.31	3.46	+1.15
Confident student writing is their own	2.62	3.54	+0.92
My classroom has clear AI norms	4.23	4.69	+0.46

75% plan to continue to have students use Byline with AI Coaching turned on.

FINDING 2

Students: decreased concern about being falsely accused of using AI — and the more they used Byline, the more they gained

Across 349 matched responses, student worry about being falsely accused of using AI showed a small but statistically significant decline, concentrated among students who actually used Byline. Values are baseline-adjusted gains in scale points (1–5 scale).

	Degree of Byline Usage		
Baseline-adjusted gain	Light	Medium	Heavy
Less worry about AI accusations	+0.06	+0.33	+0.64
Writing feels meaningful	+0.29	+0.47	+0.70
I care about improving my writing	+0.26	+0.42	+0.58

Students were not randomly assigned to usage levels, so these are associations with usage intensity, not causal effects.

FINDING 3

Students: increased engagement with writing

Across the same 349 matched responses, all changes below are statistically significant. Scores are means on a 1–5 scale.

Survey item	Pre	Post	Change
Writing feels meaningful	3.12	3.39	+0.27
I care about improving my writing	3.57	3.80	+0.23
Comfortable sharing rough drafts	3.51	3.74	+0.23
I keep going when writing gets hard	3.64	3.84	+0.20

A strong paper comes from real effort	4.14	4.23	+0.09
---------------------------------------	------	------	--------------

METHODS

How the study worked

Students and teachers completed surveys before and after the Spring 2026 pilot, answering on a 1–5 agreement scale. Responses were matched pre-to-post, and each outcome was tested with paired-samples t-tests and applying appropriate controls to make sure any overall changes were statistically meaningful rather than just random chance.¹

The analysis was conducted by Dr. Joshua Wilson, Associate Professor in the School of Education at the University of Delaware, whose research focuses on how automated writing evaluation can improve the teaching and learning of writing.

This research was funded by Tabwrite. Dr. Wilson designed and conducted the analysis.

LIMITATIONS

What this study can and can't tell you

We think research summaries should be honest about limitations, so here are ours:

- This was a single-district, single-semester pilot without random assignment — it shows real pre/post change, not proof of causation.
- The teacher sample (13 matched responses) is small; teacher results should be read as exploratory.
- All outcomes are self-reported survey responses, not writing-quality measures.
- Student effect sizes are modest in absolute terms (about 0.2–0.3 points on a 5-point scale), though consistent across a large matched sample.

Full technical reports, including item-level statistics and methodology, are available to districts on request.

¹ To address Type I error across multiple tests, we applied the Benjamini-Hochberg procedure to control the false discovery rate.