Can AI Support Writing Skill Development?

Pilot evidence from TrueMark's AI-enhanced writing platform

Quantily

This is a summary of an exploratory mixed-methods evaluation of TrueMark, an AI-enhanced writing platform for middle and high school teachers and students, piloted from February to May 2026. TrueMark's platform has two primary student-facing features, an interactive AI tutor (Copilot) and a feature that checks a student's draft against the assignment's rubric in real time, referred to as the "rubric-checker". On the teacher side, TrueMark introduced a Skills Dashboard, which estimates student proficiency levels for grade-specific writing skills, to help teachers identify where to provide targeted support. The Skills Dashboard and AI Copilot work together, informing how the system tailors supports based on individual student need.

The study was funded by the Institute of Education Sciences (IES) Small Business Innovation Research (SBIR) program, Phase IB (Award No. 91990025C0079). The primary aim of the research pilot was to understand the usability, feasibility, and potential effectiveness of a student-facing, personalized AI Copilot. Additionally, the pilot study tested an MVP of the Skills Dashboard.

01 STUDY DESIGN & DATA SOURCES

Eight teachers and their students participated in the pilot. The four participating schools were diverse both in geography (Nevada, New Jersey, and New York) and context (charter, public, and private). Qualitative data were analyzed with applied thematic analysis; quantitative analyses were correlational, including pre/post t-tests and OLS regression.

Pre/post student surveys. Administered mid-February and May 2026 with active parental consent. **(n = 50 matched)**

In-app student reflections. Optional prompts after every third assignment, with students giving repeated ratings for change scores. **(n = 25)**

Skill estimates & usage data. Continuous estimates (0–1) of proficiency levels at pilot start and end across five writing dimensions: Claim & Thesis, Evidence & Support, Analysis & Reasoning, Language & Style, Organization & Structure. **(n = 369–417 per skill)**

Teacher interviews. Conducted April–May 2026 and coded thematically, including reaction to an interactive mockup of the Skills Dashboard built on real pilot data. **(n = 8)**

- ## Writing skills grew across all five dimensions

  Skill estimates improved significantly on every dimension tracked by the platform (all $p < .001$; mean growth +0.017 to +0.024), with significant upward proficiency movement on four of five skills and no more than 2.5% of students ever moving downward in proficiency. Growth was strongest in Analysis & Reasoning and Evidence & Support.

- ## A clear dosage–response pattern

  Students who completed more assignments showed greater growth on all five skills ($r = .26–.53$, all $p < .001$), and the association held after controlling for starting skill level (+0.005 to +0.012 per additional update). Students who began at lower skill levels grew the most.

- ## Students recognized the platform's value

  Students agreed TrueMark improved their writing across all five competency areas (post-only mean = 3.87/5, $p < .001$). Two items changed significantly: that TrueMark has features that help students complete assignments (+0.38, $p = .001$) and that its feedback helps them prioritize the most important changes (+0.24, $p = .044$).

  > *"I liked how it was helping me, not giving me answers, and gave me feedback on what I need to fix."*

- ## Teachers saw student progress and changed their instruction

  Seven of eight teachers reported concrete indicators of student progress: better structural organization on handwritten assignments completed without the platform, more deliberate questions, and essay structure so internalized that students could write a well-organized essay without a checklist to follow. Six of eight also described real instructional shifts: one increased student writing volume by two to three pieces per marking period, two began grading more rigorously when students use the AI Copilot and rubric checker, and another used the platform as a formative study tool to simulate an AP testing environment.

  > *"It's embedded in their brain: if I'm writing an essay, I'm gonna have an intro, a minimum of two body paragraphs, and a conclusion. I stopped giving them checklists. They don't even need them."*

02 KEY FINDINGS, CONTINUED

- **Teachers found it usable, feasible, and worth adopting**

  All eight teachers found core features intuitive; AI-generated grading comments saved time, and detection of student AI use was the most cited driver of adoption. Teachers framed the platform as keeping student work authentically the student's.

  > *"In a classroom of 35 students, I can only help one at a time. Now every single student has access to help at the same time."*

03 AREAS FOR REFINEMENT BEFORE PHASE II

The pilot was designed to inform a Phase II proposal for a more rigorous, larger-scale study. Two priorities follow:

PRIORITY ONE

Enhance onboarding and on-demand resources

Six of eight teachers identified onboarding gaps on advanced features, asking for live demonstrations, sample class data to explore the platform with before student rollout, and an on-demand repository of help videos. Teachers also requested subject-specific training, drilldowns within skill estimates, and explanations into how proficiency levels and rubric scores are calculated.

PRIORITY TWO

Improve auto-grading accuracy

Because skill estimates derive from rubric scores, auto-grading accuracy is a prerequisite for trustworthy dashboard data. TrueMark should ensure auto-grading phantom errors are eliminated, and auto-grading is based on grade-level learning objectives. In some instances, teachers reported the system graded too harshly with lower grades and too generously with upper grades.

Taken together, the pilot offers a promising early signal of TrueMark's usability, feasibility, and effectiveness, alongside a clear set of refinements to resolve before formal outcomes testing in Phase II.