Exploring High School Teacher and Student Engagement with the Wisdom. K12 Automated Writing Evaluation Tool in the Northeastern United States: A Multiple Intrinsic Case Study

Traci C. Eshelman, Ph.D.
Liberty University
Email: TEshelman@Liberty.edu
Orchid: https://orcid.org/0000-0003-0234-8759

Abstract
The purpose of this multiple intrinsic case study was to describe how Northeastern United States middle school teachers and students engaged with a new automated writing evaluation tool used to score and provide feedback on extended essay assignments to improve teaching and learning writing. Richard Elmore's (1993) instructional core framework is the theory guiding this study. The study's central research question is: How do public and private middle school students and teachers engage with the automated writing evaluation program WisdomK12? The study leveraged an intrinsic case study design and triangulated data from educational artifacts, individual interviews, and questionnaires. Results indicated that students and teachers found WisdomK12 to save time, provide relevant, encouraging, and authentic feedback, and inspire them to write more. Implications and future research are discussed.
Keywords: writing, feedback, artificial intelligence, educational technology, automated writing evaluation, assessment, automatic assessment, text grading, extended essay assessment, automated grading

Overview

"I'm wary of having students use it to do their writing and then me use it to do the grading- just bots talking to bots" (Cardon et al., 2023, p. 272). This quote epitomizes the dilemma educators face everyday regarding when and how they should incorporate artificial intelligence (AI) into education. AI can enhance learning, but research also indicates it can inhibit learning (Cardon et al., 2023). Over the last 40 years, students' skills in writing, the second "r" in reading, writing, and arithmetic, have significantly declined (Vang et al., 2023). Teachers provide fewer opportunities for extended essays and have replaced rich and robust essay assignments with multiple-choice and short answers to save time. Consequently, students' critical thinking, reasoning, and creative skills have diminished (Giouroukakis et al., 2021). This study evaluated a program that could restore robust writing in the ELA curriculum by evaluating a new automated writing evaluation (AWE) tool called WisdomK12 at a public and parochial middle school in the northeastern United States.

Background

Writing has been an integral part of learning since the beginning of education. Over the ages, in the United States, writing was the third component of the foundations of education: reading, writing, and arithmetic. Writing improves critical thinking, logic, reasoning skills, and helps students organize thoughts, enhances ideation, improves research and revision skills (Langer, 1987). Extended writing fosters inquiry across multiple disciplines (Giouroukakis et al., 2021; Graham et al., 2018). Cross-disciplinary writing increases subject matter retention and transfer (National Commission on Writing, 2003). Writing skills improve college success, job opportunities, and advancement (National Commission on Writing, 2003). Teaching writing requires complex assignments that require reading, evaluating, and annotating articles (Vang et al., 2023). Writing may involve learning to use graphic organizers and outlines and building writing skills from short answers to longer, more intricate, and complex paragraphs and extended essays (Vang et al., 2023). Writing teaches students how to evaluate, paraphrase, and explain complex concepts and texts in simple terms and helps them structure arguments and communicate to convey knowledge (Langer, 1987; Vang, 2023). Extended essays and their assessment involving continuous feedback loops nurture deeper learning of all content in all disciplines, including history, math, and science (Gao, 2024).

Writing is an integral part of education, yet teachers have strayed from assigning extended essays over the years due to time constraints, lack of skills, and standards-driven pedagogy (Graham, 2019). High school English classes spend 6% of instructional time on writing strategies and models and 4% on evaluation and feedback (Vang, 2023). According to Graham (2019), teachers spend less than one hour daily on writing and provide no more than one or two extended writing assignments per year. Common Core State Standards (CCSS) emphasize paraphrasing and

informative text, which fosters short-answer responses. De-emphasizing extended essay writing marginalizes deep writing and critical thinking (Vang, 2023).

Langer (1997) contended that if teachers do not have confidence, education and background, and professional learning on best practices in writing, they will not have the confidence or skills to properly instruct and assess writing which in turn reduces the quantity and quality of writing instruction and evaluation. The problem is that students' writing skills are diminishing due to the lack of exposure to quality writing instruction and evaluation (Vang et al., 2023). Since so little time is now spent on extended writing, student success in college and their careers is in jeopardy. Our students' writing skills decline is evident in recent statistics from the Public Policy Institute of California (2023), which reported that 80% of all California community college students required writing remediation in 2023. According to the Center for American Progress (2023), the cost to remediate students in colleges in America is $1.3 billion. Consequently, 43 million Americans are illiterate (National Center for Education, World Atlas, 2023). This study aimed to explore and describe the responses of middle school teachers and students in the Northeastern United States regarding the implementation of WisdomK12, a new automated writing evaluation tool designed to score and provide continuous feedback on extended essay assignments.

Research Questions

The central research question addresses student and teacher experiences using WisdomK12, an AWE tool. The subsequent research questions (SRQs) delved deeper into student and teacher experiences using the tool to improve relationships, enhance writing skills, and minimize the time required to evaluate extended essays. Further, the research questions aimed to determine whether the students and teachers found the tool engaging and encouraging during the writing process.

Central Research Question

How do middle school teachers and students in the Northeastern United States describe their experiences using WisdomK12, a new automated writing evaluation tool designed to score and provide feedback on extended essay assignments?

Sub-Question One

How do middle school students and teachers in the Northeastern United States describe the quality of feedback from WisdomK12?

Sub-Question Two

How do Northeastern United States middle school students and teachers perceive their writing skills after using WisdomK12?

Sub-Question Three

How do Northeastern United States middle school students and teachers perceive student-teacher engagement after using WisdomK12?

LITERATURE REVIEW

Theoretical Framework

This study leveraged Richard Elmore's instructional core framework (ICF). Elmore extended ICF's roots in behaviorist and constructivist learning theories by focusing on the interdependent and harmonious relationships among teachers, students, and content. (City et al., 2009). The theory emphasizes student-centered learning and purports that any change in teacher, student, or content impacts one another (Elmore, 1993). ICF consists of seven pillars (Elmore, 2008). First, student learning increases with higher quality content, profound teacher subject matter knowledge, and student engagement. Second, ICF components are interdependent. Third, If components within ICF are unclear, they do not exist, and learning will not effectively or efficiently advance. Fourth, Student performance depends on the given task and how they execute these tasks. Fifth, observations and analysis drive accountability and assess whether the tasks have been completed. Sixth, students and teachers learn by doing. Finally, instruction should describe, analyze, predict, and evaluate.

Similar to Mishra and Koehler's (2006) (technological, pedagogical, and content knowledge (TPACK), ICF contends that teachers must have deep content and pedagogical knowledge to deliver lessons effectively and must be nimble to adjust to student's learning characteristics. ICF is also similar to TPACK in that each component is interdependent. The ICF, however, focuses more on the teacher-student dynamics and how challenging content can improve teaching and learning (City et al., 2009). Although TPACK infuses the technological knowledge component, this study aimed to leverage ICF to evaluate the relationship and engagement between the student,

teacher, and content when AWE becomes part of the teacher toolbox. The framework provided a lens to observe the interplay between teacher, student, and content with the advances in AWE. The framework offered guidance to understand pedagogical changes that WisdomK12 injected in writing instruction. Ultimately, this framework served as a foundation for understanding the interplay between students and teachers as they used WisdomK12 and underpinned WisdomK12's potential to improve teaching and learning through improved engagement and learning outcomes and student and teacher-improved efficacy.

Related Literature

The decline of writing education in the United States has been significantly influenced by the adoption of standards-driven curricula, such as Common Core, which have led to the decline in writing instruction. Instead of fostering critical thinking and creativity, many educators assign brief, formulaic responses, limiting students' opportunities for extended writing. Factors such as inadequate preservice teacher training and large class sizes exacerbate the issue as teachers struggle to provide substantive feedback on writing assignments. To address this decline, some educators and researchers are turning to automated writing evaluation (AWE) systems, which offer potential affordances such as fast feedback, increased student engagement, and personalized learning. However, the mixed results of AWE tools, particularly around feedback quality, accuracy, and critical thinking development, have sparked debates about their role in writing instruction. This section synthesizes the challenges and opportunities posed by AWE in reversing the current decline in writing education, considering its potential to enhance the writing process and its limitations in delivering meaningful, contextualized feedback.

Writing Education's Decline

Common core and standards-driven curricula have created the opposite of their intent of promoting critical and deeper thinking by minimizing writing assignments to multiple-choice and short answers (Vang et al., 2023). Efficacy, lack of preservice training, math/science writing training, and teacher efficacy have also hurt teachers' providing meaningful and robust extended essays (Graham, 2019). Further, how teachers approach writing is linear, limiting creativity (Benjamin & Wagner, 2021). The linear process in teaching writing excludes critical thinking, argument and persuasion, analysis, and synthesis of ideas and promotes monotony and repetition. Descriptive writing is becoming extinct. Common Core State Standards, lack of time to grade extended essays, and lack of teacher knowledge have all contributed to the decline in promoting quality writing.

Essential to quality revisions is the feedback loop (Langer, 1997). Fan and Ma (2022) contended that teachers no longer initiate a continuous feedback loop essential to revisions due to time constraints. Providing grammatical and mechanical feedback is not enough to provide quality writing instruction. Further, the feedback that studies show today gives is inconsistent, biased, and untimely (Benjamin & Wagner, 2021; Giouroukakis et al., 2021; Graham, 2019; Vang et al., 2024)

All these contribute to teacher reluctance to assign extended writing tasks and put a wedge between teachers, making learning robotic and dull, lacking critical thinking skills, minimizing creativity, marginalizing reasoning, and an overall decline in education. Evidence of the decline in writing is in the latest statistics from the Public Policy Institute of California (2023) that confirmed California community colleges had to provide remediation to 80% of students. Remediation costs in America soared to $1.3 billion per year (Center for American Progress, 2023), and 43 million Americans are illiterate (National Center for Education, World Atlas, 2023). Rather than address the problem, colleges and universities are removing writing entrance exams, and the SAT and ACT have made the writing component optional in their tests (Sorenson, 2022).

Artificial intelligence in writing

AI's application in education is not without controversy (Chen & Lin, 2024). Educators in educational technology have concluded that AI can help or hinder and offer affordances or limitations in education. The greatest hindrances in leveraging AI in education include limiting critical thinking skills and marginalizing reasoning, communications, and relationships (Al-Zahrani, 2024). Paradoxically, many of AI's affordances include these same concepts: improved relationships, communications, critical thinking, and reasoning skills (Cinque, 2024). AI in education can save educators valuable time analyzing data and personalizing learning pathways by analyzing and reporting individual students' learning styles and preferences (Tian, 2024). LLMs can evaluate entire lessons and improve pedagogy, content, and overall lessons to increase student achievement (Tian et al., 2024). AI enables educators to input complete lessons and curricula, and then the AI will evaluate and suggest improvements. Educators can even upload full transcripts of classes and have AI assess their effectiveness.

Automated Writing Evaluation

AWE is a potential solution to declining student writing skills (Cardon et al., 2023). AWE is an AI writing evaluation system based on NLP algorithms that analyzes and provides substantive feedback on grammar, syntax, style, and tone (Vang, 2023). Newer AWE programs offer customizable rubrics and robust analytics and reporting (Fagbohun et al., 2024; Gao et al., 2024). AWE is revolutionizing the teaching and learning of the writing process. In some cases, AI puts a wedge between teacher and student, but other researchers argue it can bridge the relationship between students and teachers. Some AWE tools include Grammarly, Turnitin, WisdomK12, Coh-Metrix, Mi-Writer, Google Docs, and Writing Pal (W-Pal) (Marchionda, 2023; Omid, 2022). AWEs are learning human linguistic models at an uncomprehensive pace (Fagbohun et al., 2024).

The studies' results are mixed and even conflicting. Figure 1 shows a convergence of affordances and limitations of integrating AWE in the classroom.

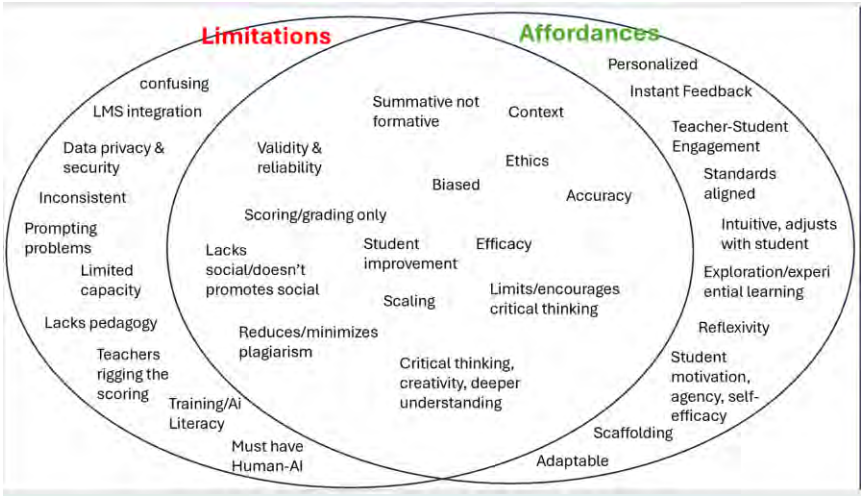

Figure 1: *AWE Limitations and Affordances*
Note. Information adapted from Gao et al. (2024).

The rapid evolution of AWE tools may explain these discrepancies. AWE is evolving at lightning speed by learning exponentially fast when earlier programs struggled to provide substantive, consistent, and accurate feedback on complex texts (Fagbohun et al., 2024). Newer language models using bidirectional transformers are more accurate and adaptable, can evaluate text based on context, not just grammar, and are becoming more efficient at analyzing longer texts.

Automated Writing Evaluation Affordances

AWE programs provide numerous affordances to educators and learners in all educational domains. All studies have noted the rapid speed at which AWE provides feedback (Fagbohun et al., 2024; Gao, 2024; Vang, 2023). One of the reasons educators stray from assigning extended essays is because of the time it takes to grade. AWE reduces the scoring time from hours to seconds. According to Vang's (2023) study, this immediate feedback increases motivation. AWE also promotes critical thinking, creativity, and a deeper understanding of the content (Gao et al., 2024; Fagbohun et al., 2024; Organnisciak, 2023). Organnisciak et al. (2023) was the only quantitative study that found AI can accurately evaluate critical thinking skills. However, this new research has not been replicated and is countered by most other studies.

Omid (2022) also contended that AWE minimizes plagiarism rather than increases it, which conflicts with the extant literature. Omid's research directly conflicts with other research that links AWE to increased plagiarism (Cardon et al., 2023).

Research is also conflicted with AWE's reliability and objectivity. Omid (2022) and al Braiki et al. (2020) asserted that AWE is reliable and objective. Tian et al. (2024) agreed that AWE was more accurate than human evaluation due to bias and low expectations. However, multiple research studies questioned AWE's accuracy and bias (Cardon et al., 2023; Fagbohun et al., 2024; Gao et al., 2024). Further researchers need to address these discrepancies. As these programs quickly evolve, results will change.

Other researchers have shown that AWE promotes reflexivity, exploration, and experiential learning, which can lead to improved writing skills (Dugartsyrenova, 2020). Research also indicates that positive and encouraging

feedback that AWE provides students can improve learning outcomes (Lin et al., 2023). Writing feedback must be substantive and encouraging for students to improve (Giouroukakis et al., 2021). However, limited quantitative research exists on whether AWE improves student writing skills. Fan and Ma's (2022) study found that when using a control group vs intervention group, the group AWE helped students more than the control group that did not use the AWE. Writing performance increased in the group that used the AWE compared to those that did not. Additionally, Fan & Ma's (2022) results contradicted each other because another group's scores were no different from the intervention group on another feedback loop. One feedback loop showed improvement with the group of students using AWE, and the other feedback loop showed no significant difference in scores. No other known studies have focused on pre-post improvements in middle school or high school student writing skills when using AWE systems.

A significant deterrent in providing robust writing assignments is scalability. Large class sizes in higher and secondary education deter teachers from providing quality and quantity writing assignments. AWE counters this with the speed and accuracy of scoring and providing feedback quickly (Fagbohun et al., 2024; Gao et al., 2024). However, Huawei's (2023) study countered Gao et al.'s (2024) argument, stating that AWE is unprepared for scaling to large classrooms.

AWE can improve student-teacher relationships because students see the machine responses as more objective than their teachers' criticisms in their feedback loop (Fagbohun et al., 2024; Tian et al., 2024). Vang's (2023) study demonstrated stronger student-teacher relationships due to increased time in evaluating AI feedback; instant and actionable feedback allowed for conferencing, revisions, critical thinking, independent learning with AI, and actionable improvements before even conferencing with the teacher, and the time teachers save in reading and manually grading provided them more time to engage and conference with the students.

AWE promotes independent and personalized learning (Fagbohun et al., 2024; Gao et al., 2024; Tian et al., 2024). Vang (2023) showed that AWE enabled students to work independently. The feedback provided empowered and encouraged students to revise their drafts independently. Vang's (2023) study showed that diagnostics helped direct writing instruction and convey the writing process to studies. The analytics guided teachers to direct instruction and encouraged specific pedagogical models like modeling to reach students better. The analytics provided enough information so teachers could leverage the correct teaching style to reach all students. Matelsky et al. (2023) agreed that automatic feedback systems provide prompt and customized feedback, increase content knowledge, and provide suggestions for improvement, which counters most other studies that say most LLMs are quick but do not offer customized feedback and do not personalize the feedback and do not offer suggestions for improvement.

Several studies demonstrated improved teacher efficacy (Gao et al., 2024; Vang, 2023). AWE taught educators the writing process and how to provide more substantive and encouraging feedback. This growth promoted teacher efficacy (Fagbohun et al., 2024). Teacher efficacy was significant with teachers with little to no experience in the writing field. Fagbohun et al. (2024) contended that AWE is a professional learning tool for teachers that trains them to grade essays, helps them recognize patterns in student writing, and uncovers common errors and misconceptions.

Automated Writing Evaluation Limitations

Depending on AI and AWE for writing can create a dependence on AI and decrease efficacy (Cardon, 2023). Depending on AI can create a loss of agency, curiosity, discovery, and motivation to learn. Many researchers contend that using AWE and AI in writing decreases critical thinking skills (Cardon et al., 2023), while other studies are mixed (Abduljabar, 2024). Cardon 2023 found that 77% of the instructors surveyed believed AI reduces critical thinking skills, and 75% believed AI minimized creativity in writing and coined the term creativity atrophy. Abduljabar's (2024) study was mixed. Some of Abduljabar's participants agreed that AI inhibited creativity and critical thinking, yet others perceived many benefits of leveraging AI in educational settings. Further research is needed to understand how AWE can enhance or hinder critical thinking and creativity.

Researchers have found AWE to be inconsistent, inaccurate, and producing hallucinations (Bang et al., 2023; de Winter et al., 2023; Fagbohun et al., 2024; Bang et al. (2023) found that ChatGPT was only 64% accurate in 10 different categories of reasoning and is inconsistent in reasoning producing mixed results in writing. Further, research has indicated that AWE cannot provide robust, tailored, and contextualized feedback (Huawei et al., 2023; Palermo & Wilson, 2020; Zhu et al., 2020). These studies contradict many claims that AWE is consistent, accurate, and can provide robust, tailored, and contextualized feedback (al Braiki et al., 2020; Omid, 2022; Tian et al., 2024). Researchers must conduct further studies to address these discrepancies.

AWE and AI in writing are causing significant teacher anxiety (Huawei & Aryadoust, 2023). Cardon et al. (2023) noted that teachers fear redundancy and feel overwhelmed by new technology's rapid development, evolution, and demands. Teachers lack training on how to use AWE and lack AI literacy and AI pedagogy, which would support integrating AWE systems (Cardon et al., 2023; Omid, 2022). These stressors have created a negative attitude and decreased teacher efficacy and motivation (Omid, 2022).

Bias, ethics, data privacy, and security are all concerns that Fagbohun et al.'s (2024) study addressed. Depersonalization is another issue that concerns researchers (Fagbohun et al., 2024; Fischer & Hagel, 2024). Other limitations researchers noted include AWE's inability to comprehend students' cognitive skills (Fischer & Hagel, 2024), its lack of robust and deep contextual feedback and representational thinking (Fischer & Hagel, 2024; Gao et al., 2024; Huawei & Aryadoust, 2023; Zhu et al., 2020), and its limitations to scoring and grading rather than providing robust feedback and specific examples to improve writing and organization (Fischer & Hagel, 2024; Gao et al., 2024). AWE platforms are typically summative rather than formative, according to Winter et al. (2023), who agreed and added that they only offer grammatical and mechanical feedback. Finally, Huawei and Aryadoust (2023) concluded that AWE discourages students because it takes the social aspect out of the writing process.

Summary

The extant literature provides a solid foundation for how teachers use AI in writing (Cardon et al., 2023; Omid, 2022) and provides a variety of AWE tools that are currently on the market (Omid, 2022; Marchionda, 2023). Recent literature has provided an extensive quantitative review of AWE's efficacies, limitations, and deficiencies, but few qualitative studies have provided a deeper and richer understanding of educators' experiences with AWE (Gao et al., 2024). Research has clearly stated the importance of writing to promote critical thinking, creativity, and reasoning (Graham et al., 2018). Over the past forty years, America has experienced a decline in quality writing education, and Common Core State Standards have exacerbated this decline (Vang et al., 2023). Multiple gaps in the research remain. Extant literature has revealed multiple systematic literature reviews (Al Braiki et al., 2020; Fan & Ma, 2022; Gao et al., 2024; Huawei & Aryadoust, 2022; Omid, 2022), but few studies evaluated AWE and analyzed how it works and how teachers responded to using it. Most studies were quantitative, and few offered qualitative perspectives (Tian et al., 2024). After thoroughly evaluating the literature, no known peer-reviewed study has evaluated student feedback on using any AWE tool. Recent research has neglected to describe and evaluate how teachers and students engage with software program architecture (Fischer & Hagel, 2024). Gaps in high school and middle school essay evaluation (Latif & Zhai, 2023) and limited research evaluating domain-specific essay assessment exist (Fagbohun et al., 2024; Fischer & Hagel, 2024). Fagbohun (2024) also highlighted critical gaps in evaluating LLM formative feedback, understanding how AWE supports personalized learning, adapting assessment methods to individual student learning plans, and identifying cognitive patterns to improve personalized interventions. This study aims to qualitatively analyze how teachers and students perceive the latest AWE technology integrated into private and public Northeastern United States middle schools, which will address these gaps.

Research Design and Methodology

This study employed a multiple case study design by leveraging artifacts, interviews, and questionnaires to determine student and teacher perceptions of the AWE program WisdomK12. The aim was to understand how WisdomK12 provided feedback to students and how they interacted with the program to improve writing through continuous feedback loops. Data analysis provided thematic descriptions of how students and teachers responded to the substantive feedback loops WisdomK12 provided during multiple extended essay prompts during the spring semester 2024 (February – May 2024). Data analysis involved providing a detailed description of each case, thematic analysis across each case, and finally, making assertions regarding the cases (Stake, 1995). Triangulation of data from the artifacts, interviews, and questionnaires provided the basis for synthesis and direct interpretations led to developing naturalistic generalizations of the cases (Stake, 1995).

Setting and Participants

This study used two middle schools located in the Northeastern United States. The first was middle school, which was a public school, and the second was a parochial school. Each school's population was mainly white or Caucasian students of middle-class socioeconomic status. Both schools reported increased behavior and attitude problems due to the recent COVID-19 shutdowns, and teachers noted significant differences in student motivation and attitude compared to pre-COVID-19 effects. Purposeful sampling provided four teachers and 47 students. The parochial school provided 27 seventh-grade students, and the public school provided 20 eighth-grade students. The parochial school had one English Language Arts (ELA) teacher participate, and the public school had three ELA teachers participate. Consent and assent were obtained at each location.

Procedures

From October through May 2024, ELA teachers from two Northeastern United States public and private schools assigned their typical writing essays and prompts throughout two semesters. Teachers applied WISDOMK12 and graded manually to compare findings. Both public and private schools applied evidenced- and standards-based writing rubrics. The classes used

Teachers provided writing prompts and feedback based on the semester's assignments, loaded the content into WidsomK12, and gave students a login to submit the assignment through WisdomK12. Teachers evaluated the AWE's feedback before giving the AI-generated feedback to the students to ensure accuracy and understanding. Upon completion, data was collected from the WisdomK12 feedback loops, questionnaires, and individual and semi-structured interviews with all students and teachers participating in the study.

Private School Procedures

The private school teacher used WISDOMK12 in two domains: ELA and technology courses. In October 2023, the teacher first introduced WISDOMK12 to two eighth-grade ELA classrooms in a private school in the Northeastern United States. This eighth-grade class was significantly lower academically than any class the teacher had taught in the previous 48 years. These students were the class most affected by the pandemic and negatively impacted by social media. Students had done little writing in sixth and seventh grades.

Chromebooks were not available when the school year started. Instead of using WISDOMK12 SWIPES so that both the teacher and AI could analyze baseline writing samples, the teacher required the students to write a two-page letter of introduction. The teacher noted trends among the essays and presented about six weeks of writing instruction, often using "short writes" to help students feel successful.

The first complete assignment in WISDOMK12 was a narrative essay with a 100-pt. The rubric was modeled after the PA State Writing Rubric, cross-referenced with the NYSED Writing Rubrics. Pre-writing on paper was required for a minimum of ten minutes. The teacher circulated the room and helped the students process the personalized AI advice. After two drafts, students could copy and paste their revised essays into a Google doc, apply more "FINAL WISDOM" from the AI feedback, use the grammar and spell check, and submit for a final grade. The teacher scored the final submissions by hand but compared the WISDOMK12 final draft scores to affirm the tool's efficacy.

The second complete assignment in WISDOMK12 for this class was a choice among three persuasive essays. Topics include school cell phone usage, physical education requirements, and community laws. The teacher applied SWIPES to the assignments for students who had finished their final essays and changed the rubric in WISDOMK12 to align with the assignment.

The third assignment for which WISDOMK12 drafts were required was another narrative, "A Change of Heart." In April, the teacher used a previously released NYSED reading prompt with the RACE rubric to practice answering test questions. Several schedule changes and Chromebook conflicts prevented the students from completing the assignment before testing.

The other students who applied to WISDOMK12 were sixth—and seventh-grade students in technology classes. Using WISDOMK12 in the technology class aimed to integrate writing into a STEM subject. The emphasis was on describing technology and informative writing rather than teaching the writing process. The teacher did not use WISDOMK12 scores as part of their grade. The objective was to increase cross-disciplinary writing.

Public School Procedures

Integrating the WisdomK12 platform into seventh-grade social studies classes followed a detailed and deliberate process designed to enhance student engagement with writing assignments while incorporating technology into instruction and assessment. The process began by reviewing the curriculum and identifying reading assignments aligned with specific content areas. Teachers then crafted questions with clear answers that allowed for further elaboration, encouraging students to develop their responses fully. Students began by outlining their essays using a graphic organizer before drafting their work directly into the WisdomK12 platform. The platform provided feedback designed to help students reflect on their writing, make necessary revisions, and deepen their understanding of the material.

WISDOMK12 provided teachers with comprehensive training focused on the tool's functionality for successful implementation. This training emphasized creating assignments, using the RACE rubric (Restate, Answer, Cite,

Explain), and integrating the tool into grading. To ensure consistency between the teachers' evaluations and the platform's feedback, teachers compared their scores using the RACE rubric to those generated by the WisdomK12 AI. This comparison was conducted to assess inter-rater reliability and ensure that both human and AI evaluations were aligned.

The writing assignments themselves were based on the RACE rubric. The district established a structured approach based on evidence-based writing best practices. The teachers required students to complete these tasks three times over nine weeks. In October 2023, teachers began integrating these assignments into WisdomK12. Students submitted their work directly through WisdomK12, which applied the RACE rubric to assess their writing and provided more detailed feedback, including style and substance suggestions for revision. WISDOMK12 included suggestions on grammar, coherence, structure, and style, offering students valuable insights to improve the students' writing. The platform's feedback system also gave teachers a clearer view of student performance, allowing for more informed instruction.

During the initial implementation phase from October to December 2023, teachers used both manual grading with the RACE rubric and the AI-generated scores from WisdomK12. This process allowed for ongoing assessment of inter-rater reliability and ensured alignment between human and AI evaluations. From January onward, teachers exclusively used WisdomK12 for grading, continuing to rely on the RACE rubric for core assessments. In one approach, teachers used the feedback generated by WisdomK12 to guide students in improving future assignments. Another approach encouraged students to engage directly with feedback to revise their current work, fostering improvement through reflection and revision.

Teachers also conducted one-on-one conferences with students to discuss the detailed feedback provided by WisdomK12. These conferences were instrumental in guiding students either in revising their current assignments or preparing for future writing tasks. By focusing on targeted feedback, the conferences reinforced student learning and promoted continuous improvement in writing.

Data Analysis

The researcher analyzed the individual interviews by transcribing, coding, conducting a thematic analysis, and then interpreting the findings according to Stake's (1995) best practices. Similarly, the researcher collected and organized the continuous feedback loop data, identified patterns, conducted a contextual analysis, and iteratively reviewed the data. Finally, the researcher cleaned the data from the questionnaires, conducted a descriptive quantitative and qualitative analysis, and compared findings across each case. Leveraging a robust qualitative data analysis software program, Atlas.ti, the researcher triangulated the data by cross-referencing, evaluating frequencies, synthesizing shared insights, and finally drawing conclusions from the combined data sets based on Elmore's (1993) ICF and RQs.

Ethical Considerations

The researcher has no financial conflicts of interest, professional affiliations, or ethical complications that could compromise this research's integrity, objectivity, or validity. All study aspects were conducted independently, adhering to ethical guidelines and standards. The researcher is Collaborative Institutional Training Initiative (CITI) certified and adhered to all human ethics requirements and regulations, including respect for humans through informed consent and assent, beneficence by minimizing harm and maximizing benefits to participants, and justice by ensuring fair and equitable research practices.

FINDINGS

Participants included 27 students in the seventh grade at a private school in Northeastern United States and 20 eighth-grade students at a public school in Northeastern United States. Results were shared with all eighth-grade teachers in the public school as they prepared to implement the WISDOMK12 tool with fidelity during the 2024-25 school year. In a follow-up survey, the public school provided 18 teacher participants, and the private school had one teacher participate. The students in both schools were predominantly Caucasian from middle-class families. Most public school teachers were white females ranging from three years of teaching experience to more than 30 years of experience. The private school teacher participant had more than 40 years of experience. The students all have been experiencing extreme behavior issues and challenges in learning due to COVID-19's social and emotional impact. The private school teacher explained, "This was the most difficult group of students... Completely different creatures from my other 48 years. It (this year) was a real struggle," another teacher from the public school added. "They just bonded over being bad. And when you punish them, they wear it like a badge of honor."

Most students and teachers (n=47) were satisfied with using WisdomK12 (see Figure 2).

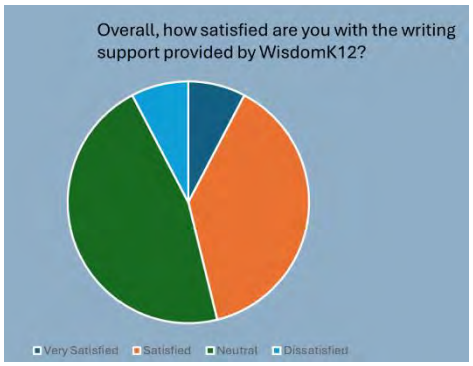

Figure 2: O*verall WisdomK12 Satisfaction*

Every participant surveyed noted improved general feedback. Most who expressed satisfaction noted improved clarity, creativity, grammar, punctuation, and general feedback over feedback teachers provided manually. Students were impressed with the instant feedback, accuracy, and ease of interpretation and application. Teachers stressed how the students valued the computer feedback over their feedback, owing to this phenomenon of student perceptions of teacher bias. All participants, except for the two outliers, believed the program provided objective, timely, accurate, and encouraging feedback. However, even the outliers who expressed dissatisfaction with the program noted on the survey that the feedback WisdomK12 provided was improved over manual feedback provided by teachers. Table 1 provides an overview of the themes generated from this study.

Table 1: *Themes & Subthemes*

Theme	Subthemes	
Personalized learning	Intuitive/Grows with student	Skills growth
Instant Feedback	Accurate	Encouraging
Relationships	Objective	Time

Personalized Learning

All teachers and students concurred that WisdomK12 tailored the feedback to suit the students' grades, literacy, and writing objectives. Bobby, from the parochial school, stated, "I love how personalized and specific it (the feedback) is." Teacher 1 from the parochial school noted that WisdomK12's feedback grew with the student's skills. "WisdomK12 adapted the language to reflect the students' skills and language abilities." Libby was surprised that WisdomK12 listened to her, stating, "I like that it actually reads my essays and pulls actual sentences out from my writing and makes them better and gives them more detail." The sentiment was that the program was listening to her and heard her voice.

Teachers were also encouraged to see how WisdomK12 provided an alternative way to reach students more personalized and directly. Several teachers noted that technology has changed the way students think and learn, even citing Johns Hopkins studies on student brain scans that indicate students are learning differently than before cell phones and personal devices were introduced. One teacher noted, "So my point is we've discovered how to reach them. Perhaps Wisdom K12 is onto something because I think some of the data showed that the kids are responding more to the AI than they are to the human."

Teachers noted how the AWE program provided personalized feedback for every type of writing assignment from persuasive to narrative. The feedback was encouraging and provided examples for improvement. The following demonstrated encouragement in the feedback regarding a narrative a student submitted in an essay on his experience with basketball, "Your enthusiasm for basketball shines brightly through your essay, making it both heartwarming and inspiring." WisdomK12 also provided examples to help students improve their writing while encouraging and motivating them to improve their writing, "You've done a great job of capturing the emotions of various moments but delve deeper. For example, when you mentioned having a joyful attitude, describe the intensity and the environment." Another essential part of the writing process that WisdomK12 addressed was reflection. Each assessment allowed the student to reflect on the writing process. WisdomK12 asked one student, upon completion, "How did your perspectives change after writing this essay?" The teachers agreed that this

encouragement, examples, and reflection pieces of the assessment opened doors to communication and deeper thinking and revising, which the teachers did not believe they could effectively and efficiently do manually.

Teachers noted that WisdomK12 was intuitive and progressed at the students' pace, thus supporting personalized learning and providing encouraging feedback. Overall, these attributes enhanced the writing process. One teacher noted that teachers no longer adhere to the writing process in her observations. This teacher noted that most English teachers never write beyond whatever they did in college or grad school, "They never wrote a thing, and they're teaching writing." Another teacher expanded, "Many skip pre-writing, which to me, I learned late in life, is the most valuable part. Just get your ideas down in some kind of web or organized form or number them and so many teachers don't even touch that." Other teachers agreed, noting that teachers skip doing revisions and neglect resubmitting the papers. Multiple teachers only submitted drafts and did not go through the revision process at all. One teacher explained, "But most teachers aren't gonna do that because they don't have the time…" Another teacher added that peer review was ineffective: "It's the blind leading the blind." The teacher added that WisdomK12 improved these inefficiencies and deficiencies, stating, "This tool solves so many of the challenges, not just the time it saves from taking those first drafts home… You can actually do the entire writing process with this tool."

Teachers believed implementing WisdomK12 improved students' writing skills. Figure 3 shows gains in student achievement after using WisdomK12 supporting the teachers' observations.

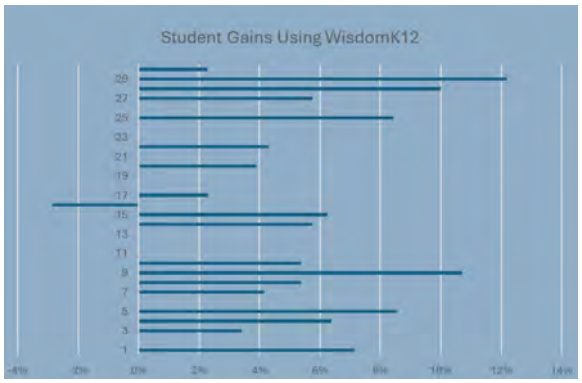

Figure 3: *Student Perceived Gains*

Instant feedback

Students and teachers from both schools concurred that the most significant benefit of WisdomK12 was receiving feedback instantaneously. Figure 4 compares the time teachers reported it would take them to grade the assignments if they had graded manually to when it took WisdomK12 to grade.

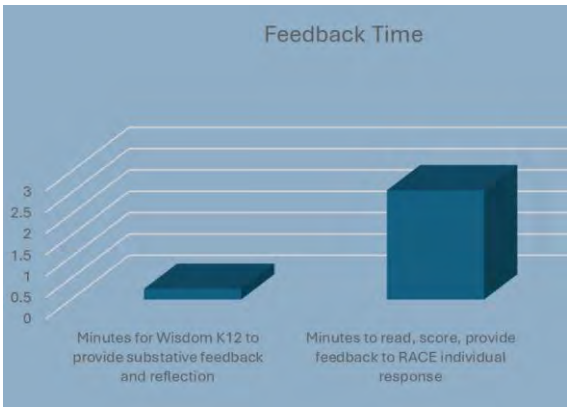

Figure 4: *Feedback Time*

Grading and providing feedback are two different tasks. "Providing substantive feedback takes substantially longer than simply grading a paper," stated one teacher. This table may not be representative of grading and scoring versus providing substantive, encouraging feedback that students can use to revise their drafts. Further research is needed to distinguish scoring from feedback. The students and teachers agreed that the feedback was accurate, unbiased, encouraging, and helpful. Abbie, an eighth-grade student, elaborated, "I can get my feedback in under a minute."

Tiffany agreed, adding, "It makes my writing more engaging for the reader rather than boring and bland." Hannah concurred, "WisdomK12 has helped me become a better writer, and it helps me make fewer mistakes and add more details." Teachers agreed that the feedback saved them time grading papers and enabled students to receive feedback for immediate revisions. According to the parochial teacher, when students have to wait an average of two weeks, they have already moved on and forgotten what they had written. The teacher explained, "When the student receives the feedback instantaneously, they can immediately apply it. It is fresh." Most students and teachers concurred that the program improved clarity and coherence and increased their ability to incorporate feedback effectively. The private school teacher explained that the students were first copying and pasting content from the draft to the final paper, but as they worked more with WisdomK12 and saw what it could do, their motivation changed. The private school teacher elaborated, "As time went on, they became more appreciative of it, and at the end, they had something to do for their religion teacher, a moral autobiography, and they were begging me to put it into Wisdom so that they could get the feedback at home."

Relationships

A critical aspect of the findings included the relationships WisdomK12 fostered between students and teachers. Students accepted the feedback from WisdomK12 without question or attitude. When receiving feedback from their teachers, the private school teacher stated, Children feel like they are being picked on or attacked." Another teacher noted, "When they received the feedback from WisdomK12, they accepted it as objective and non-threatening." This objectivity fostered and nurtured more positive relationships between the students and teachers. The time they have also empowered teachers to delve deeper into creativity, critical thinking, and multiple revisions and iterations to improve the writing because the program was doing tedious work. This allowed more time to conference with the students and build those critical relationships.

Comparison between cases

Results varied between the private and public schools. Both schools' participants agreed that WisdomK12 provided instantaneous and helpful feedback, but some findings diverged. On average, the private school's perceived confidence in writing was more significant than the public school's. See Figure 5.

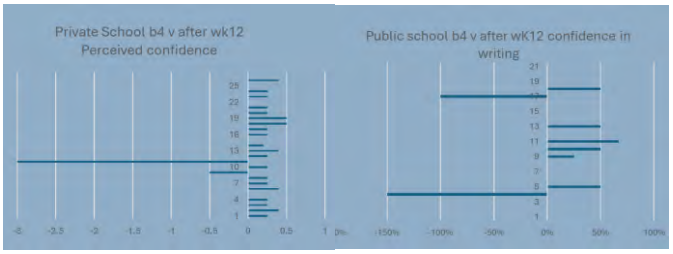

Figure 5: *Confidence Comparison Between Public and Private Schools*

The private school also indicated that significantly more students found it easier to understand and apply WisdomK12 feedback. See Figure 6.

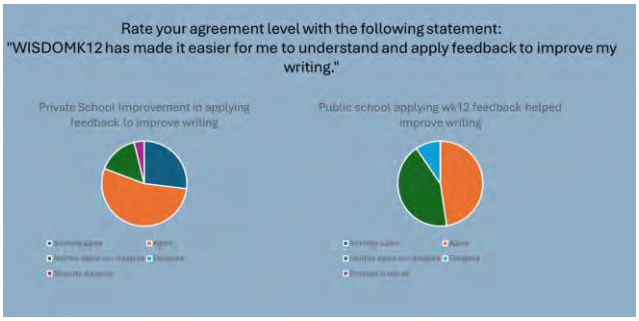

Figure 6: *WisdomK12's Feedback Clarity*

The quality of feedback, according to the participants of both schools, was comparable. See Figure 7.

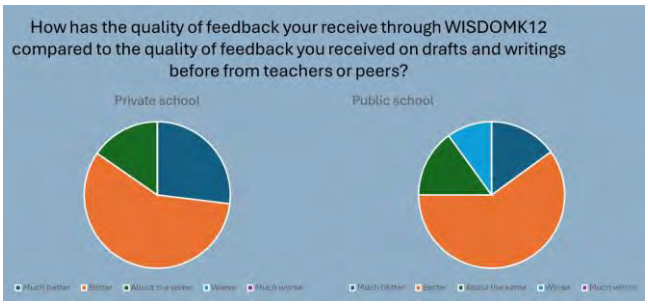

Figure 7: *Perceived Quality of Feedback*

Finally, the private school participants felt more strongly that WisdomK12 improved their editing skills. See Figure 8.

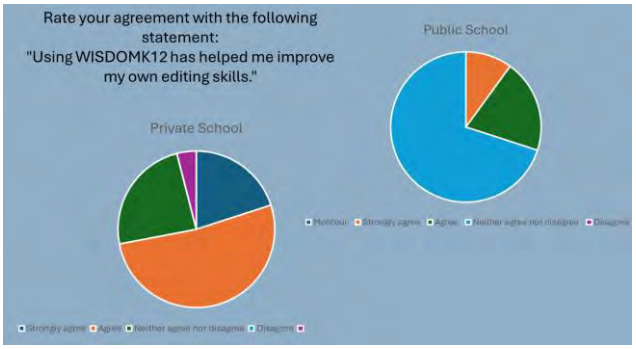

Figure 8. *Students' Perceived Improvement in Writing Skills*

Outlier Data and Findings

Two outliers existed in the findings. Two parochial and public school students gave poor ratings on the questions asked regarding WisdomK12's efficacy, user experience, and perceived skills development. Teachers explained that behavior has been an issue since COVID-19, and attitudes have been poor. The teachers believed these students disengaged and were defiant in using the program, which reflected how they behaved in all other classes. Further research is needed to explain these outliers.

Discussion

This study aimed to determine how middle school teachers and students in the Northeastern United States describe their experiences using WisdomK12. The study supported and extended Elmore's (1993) ICF theory, confirming and refuting the extant literature. Only two participants found WisdomK12 intuitive, helpful, speedy, accurate, straightforward, and encouraging. This section will discuss the findings related to the extant literature and the theoretical framework. Implications and future research will also be discussed.

Summary of Thematic Findings

Theoretically, this study demonstrated and extended Elmore's ICF (1993). WisdomK12 enhanced the teacher-student relationship and content knowledge. After triangulation, surveys, archival data, and interviews concurred that WisdomK12 increased teacher knowledge and skills in the writing process. The AWE program helped improve student-teacher relationships by providing more time and objective feedback. WisdomK12 also improved students' attitudes and motivation by personalizing the feedback and creating a bias-free environment, promoting independent learning and positivity.

The extant literature on AWE is mixed. This study affirmed AWE's affordances and negated the limitations in the extant literature. This study affirmed that AWE provides instant feedback, thus streamlining the writing process and improving revisions. Teacher-student engagement and relationships improved. Teachers found the program reliable, reflexive, and promotes student-centered learning. The teachers also believed WisdomK12 was accurate and authentic and promoted critical thinking and creativity. The students found WisdomK12 to be objective, fast, and transparent. Both students and teachers found the program overall to be satisfactory.

WisdomK12 promoted personalized learning by providing tailored feedback and specific suggestions (Fagbohun et al., 2024; Tian et al., 2024). The program tailored feedback and suggestions to learners' styles, characteristics, demographics, and preferences, thus promoting student-centered learning. Instant feedback and student-centered

learning increased engagement, participation, communication, and collaboration (Omid, 2022; Vang et al., 2023). This led to closer, more meaningful, and trusting relationships between students and teachers. The students used and applied WisdomK12's feedback more readily than they would have applied their teacher's feedback. Further, the relationships improved because the students perceived the AWE as objective and non-threatening. The AWE's objectivity encouraged and motivated the students (Lin et al., 2023; Tian et al., 2024).

This AWE program significantly improved the writing process by providing teachers and students with comprehensive, meaningful, and encouraging feedback and specific examples of how to improve their essays (Fagbohun et al., 2024; Fan & Ma, 2022; Vang et al., 2023). This feedback improved teachers' pedagogical practices and may have contributed to student gains in writing scores (Fan & Ma, 2022).

This study demonstrated that WisdomK12 fostered creativity, originality, and critical thinking by providing contextualized, consistent, and insightful feedback (Cardon et al., 2023; Fagbohun et al., 2024; Matelsky et al., 2023). Jiang et al. (2023), Cardon et al. (2023), and Palermo and Wilson (2020) research questioned AWEs' ability to provide examples and meaningful and contextualized feedback. This study refuted these findings and demonstrated that WisdomK12 provided appropriate, humanized, clear, concise feedback and specific examples. Teachers confirmed that the program fostered creativity and originality and promoted critical thinking (Fan & Ma, 2022; Organnisciak et al., 2024)

Multiple studies have criticized AWEs' stating that the programs limit critical thinking, authenticity, and creativity (Cardon et al., 2023; Gao et al., 2024). Many studies have also professed AWE's limitations in providing contextualized, conceptual, and representational thinking feedback (Cardon et al., 2023; Fisher et al., 2024; Gao et al., 2024). Fisher et al. (2024) proclaimed that teacher manual feedback was still superior to AWE, while other studies recommended cooperation between AI and humans (Cardon et al., 2023; Huawei et al., 2023; Tian et al., 2024). This study presented evidence from teachers and students that AWE-generated feedback was superior and preferred to human feedback. However, further research must be conducted to support and generalize these findings. This study, however, clearly demonstrated that WisdomK12's feedback was contextual, robust, student-centered, and intuitive. The study negated de Winter et al.'s (2023) contention that AWE's rubrics were inaccurate and ineffective with WisdomK12's robust and adjustable rubrics that can be tailored to each writing assignment and are standards-aligned.

According to de Winter et al. (2023), AWEs' feedback was inconsistent, confusing, and time-consuming, creating more work for the teacher. This study demonstrated the opposite, with most students and teachers confirming that WisdomK12's feedback was clear and easy to understand, interpret, and apply. Zhu et al.'s (2020) study showed no significant gains in students' writing skills, yet this study showed modest gains. Further research is needed to generalize and confirm these initial results.

Omid (2022) noted teachers' reluctance and negative attitudes toward using AWE. However, this study showed that teachers who learned about other teachers using the program asked for their assignments to be run through WisdomK12 throughout the year. This study showed that WisdomK12 increased teacher motivation to use AWE.

Limitations and Delimitations

This study's limitations and delimitations should guide future research. The nature of this study did not allow for generalization and was limited to one specific region and age group of the United States. Each school had slightly different procedures and integrated WisdomK12 differently, so the comparison between the public and private schools is questionable. Too many variables exist to come to a solid conclusion on the differences and similarities between public and private schools. Further, only one teacher in the private school participated, while 27 teachers in the public school participated, thus providing unreliable results.

Recommendations for Future Research

Future research must account for the abovementioned limitations by consistently applying the intervention with consistent procedures for each location. Future research should also quantitatively address the accuracy of WisdomK12. Future research should address whether an effective AWE program motivates teachers to assign more extended essays throughout the year and if AWE quantitatively increases students' writing skills. Future research should also evaluate the consistency and reliability of the feedback. Finally, research should address AI literacy and how teachers incorporate AWE into assignments across multiple disciplines.

Conclusion

Teachers and school districts have progressively deemphasized writing in the curriculum. Student creativity, critical thinking, and deep content learning have suffered because of this pedagogical shift. AWE could provide a solution to empower teachers to assign more extended essays and promote writing in the classroom again, thus improving creative thinking and creativity. This study demonstrated that a new, cutting-edge AWE program is ticking all the boxes of providing quality writing feedback and refuting past studies criticizing AWE feedback. Further research is needed to extend this study. However, the initial findings are promising and suggest that this latest AWE can provide robust, contextualized, personalized feedback, save teachers time, foster positive student-teacher relationships, minimize bias, and provide specific and tailored examples to improve student writing in middle school.

References

Abduljaber, M. F. (2024). Perceived Influence of Artificial Intelligence on Educational Leadership's Decision-Making, Teaching, and Learning Outcomes: A Transcendental Phenomenological Study.

Al Braiki, B., Harous, S., Zaki, N., & Alnajjar, F. (2020). Artificial intelligence in education and assessment methods. *Bulletin of Electrical Engineering and Informatics, 9*(5), 1998-2007.

Al-Zahrani, A. M. (2024). Unveiling the shadows: Beyond the hype of AI in education. *Heliyon, 10*(9).

Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly, 86*(3), 257-295.

Chen, J. J., & Lin, J. C. (2024). Artificial intelligence as a double-edged sword: Wielding the POWER principles to maximize its positive effects and minimize its adverse effects. *Contemporary Issues in Early Childhood, 25*(1), 146-153.

Cinque, T. (2024). Exploring the paradox: Perceptions of AI in higher education–a study of hype and scandal. *Explorations in Media Ecology, 23*(2), 199–215.

City, E. A., Elmore, R. F., Fiarman, S. E., & Teitel, L. (2009). *Instructional rounds in education* (Vol. 30). Cambridge, MA: Harvard Education Press.

de Winter, J. C., Dodou, D., & Stienen, A. H. (2023, November). ChatGPT in education: Empowering educators through methods for recognition and assessment. *In Informatics, 10*(4)

Dugartsyrenova, V. A. (2020). Supporting genre instruction with an online academic writing tutor: Insights from novice L2 writers. *Journal of English for Academic Purposes, 44*, 100830.

Elmore, R. F. (1993). *The role of local school districts in instructional improvement. Designing coherent education policy: Improving the system*, pp. 96–124.

Elmore, R. F. (1995). Teaching, learning, and school organization: Principles of practice and the regularities of schooling. *Educational Administration Quarterly, 31*(3), 355–374.

Elmore, R. (2008). *Improving the instructional core*. Draft manuscript.

Fagbohun, O., Iduwe, N. P., Abdullahi, M., Ifaturoti, A., & Nwanna, O. M. (2024). Beyond traditional assessment: Exploring the impact of large language models on grading practices. *Journal of Artificial Intelligence and Machine Learning & Data Science, 2*(1), 1–8.

Fan, N., & Ma, Y. (2022). The Effects of Automated Writing Evaluation (AWE) Feedback on Students' English Writing Quality: A Systematic Literature Review. *Language Teaching Research Quarterly, 28*, 53-73.

Fischer, D., & Hagel, G. (2024, May). *Enhancing NLP-Based Educational Assessment: A Node-Based Graph Approach for Analyzing Freeform Student Texts*. In 2024 47th MIPRO ICT and Electronics Convention (MIPRO) (pp. 108-113). IEEE.

Gao, R., Merzdorf, H. E., Anwar, S., Hipwell, M. C., & Srinivasa, A. (2024). Automatic assessment of text-based responses in post-secondary education: A systematic review. *Computers and Education: Artificial Intelligence*, 100206.

Giouroukakis, V., Kramer-Vida, L., Siegel, K., & Jaccarino, V. (2021). Addressing the High School and College Writing Gap: Democratizing Writing Instruction. *The English Record, 30*(2).

Graham, S. (2019). Changing how writing is taught. Review of Research in Education, 43(1), 277–303.Huawei, S., & Aryadoust, V. (2023). A systematic review of automated writing evaluation systems. *Education and Information Technologies, 28*(1), 771–795.

Huawei, S., & Aryadoust, V. (2023). A systematic review of automated writing evaluation systems. *Education and Information Technologies, 28*(1), 771-795.

Langer, J. A., & Applebee, A. N. (1987). *How Writing Shapes Thinking: A Study of Teaching and Learning*. NCTE Research Report No. 22. National Council of Teachers of English, 1111 Kenyon Rd., Urbana, IL 61801 (Stock No. 21802-222, $6.95 member, $8.95 nonmember).

Latif, E., & Zhai, X. (2024). Fine-tuning ChatGPT for automatic scoring. Computers and Education: *Artificial Intelligence, 6*, 100210.

Lin, J., Thomas, D. R., Han, F., Gupta, S., Tan, W., Nguyen, N. D., & Koedinger, K. R. (2023). *Using large language models to provide explanatory feedback to human tutors*. arXiv preprint arXiv:2306.15498.

Marchionda, M. (2023). *From Chalkboards to ChatGPT: Harnessing the Power of AI in the Classroom*. Wisdom Media Group: Pittsburgh, PA.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017-1054.

Meyer, J. G., Urbanowicz, R. J., Martin, P. C., O'Connor, K., Li, R., Peng, P. C., ... & Moore, J. H. (2023). ChatGPT and large language models in academia: Opportunities and challenges. *BioData Mining, 16*(1), 20.

National Commission on Writing. (2003, April) *The neglected R: The need for a writing revolution*. https://archive.nwp.org/cs/public/print/resource/2523

Omid, M. (2022). Review of research on the use of Information and Communication Technologies (ICTs) in ELT-related academic writing classrooms. *Journal of Language and Education, 8*(2 (30)), 165–178.

Palermo, C., & Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: A mixed-methods study. *Journal of Writing Research, 12*(1), 63-108.

Stake, R. (1995). *The art of case study research*. Sage.

Sorenson, T. (2022, March 28). *The Optional SAT Essay: What to Know: Tackling this section of the SAT requires preparation and can boost some students' college applications.* U.S. News and World Report. https://www.usnews.com/education/blogs/college-admissions-playbook/articles/what-to-know-about-the-optional-sat-essay.

Tian, Z., Sun, M., Liu, A., Sarkar, S., & Liu, J. (2024). *Enhancing Instructional Quality: Leveraging Computer-Assisted Textual Analysis to Generate In-Depth Insights from Educational Artifacts*. arXiv preprint arXiv:2403.03920.

Vang, M. C. (2023). *A Descriptive Analysis of Teachers' Writing Instruction Using Project Topeka, An Automated Essay Scoring Tool*. Executive summary. digitalpromise.org

Zhu, M., Liu, O. L., & Lee, H. S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, p. 143, 103668.