Received: 1 August 2024 | Revised: 8 November 2024 | Accepted: 11 November 2024 | Published online: 18 November 2024

REVIEW

Recommendations for Integrating Automated Writing Evaluation with Evidence-Based Instructional Practices

International Journal of Changes in Education
yyyy, Vol. XX(XX)
DOI: 10.47852/bonviewIJCE42024011

BON VIEW PUBLISHING

Joshua Wilson[1,*], Tania Cruz Cordero[1], Andrew Potter[2], Matthew Myers[1], Charles A. MacArthur[1], Gaysha Beard[3], Emily A. Fudge[3], Alexandria Raiche[4] and Cristina Ahrendt[5]

1 School of Education, University of Delaware, USA
2 Department of Psychology, Arizona State University, USA
3 Red Clay Consolidated School District, USA
4 Milford School District, USA
5 University of Delaware

***Corresponding author:** Joshua Wilson, School of Education, University of Delaware, USA. Email: joshwils@udel.edu

Abstract: Automated writing evaluation systems are formative assessment systems that provide immediate, automated feedback on L1, L2, and EFL students' writing in the form of writing-quality scores and suggestions for revising. As such, these systems have the potential for alleviating some of the persistent barriers teachers face to implementing evidence-based writing instruction practices. However, simply adopting this technology without careful attention to how it is implemented will not guarantee instructional benefits. In this article we draw on prior research to make recommendations to effectively integrate automated writing evaluation alongside evidence-based writing instruction practices to improve writing instruction and intervention, leveraging the affordances of this technology while addressing its limitations. Specifically, we discuss how researchers, interventionists, and educators using automated writing evaluation should develop students' knowledge of underlying evaluation criteria; teach strategies for planning, drafting, and revising; supplement automated feedback with effective teacher-provided feedback; and enact goal setting and progress monitoring.

Keywords: automated writing evaluation, automated scoring, writing, writing instruction, writing assessment, automated feedback

1.Introduction

Writing is a cognitive and social process, wherein writers leverage cognitive and metacognitive resources such as knowledge, attention, self-regulation, and composing-specific processes and strategies, as well as affective resources relating to motivation and self-efficacy, to compose for different purposes within specific communities [1]. Writing develops, in part, from effective instruction and frequent opportunities to practice coupled with timely feedback.

Although there are different ways of defining 'effective instruction,' evidence-based practices refer to instructional methods that have been rigorously evaluated and shown through research to significantly improve learning outcomes. In the context of writing instruction, a meta-analytic review of prior meta-analyses of research on K–12 approaches to writing instruction and intervention [2] identified several evidence-based practices for writing, including ensuring students frequently practice writing, creating supportive writing environments, and enhancing students' strategies, knowledge, and motivation. Grounded in Graham's [1] Writer(s)-Within-Community model of writing development, these instructional practices support cognitive and metacognitive processes critical for writing. They activate cognitive resources like knowledge and attention, and metacognitive skills such as self-regulation and goal setting, helping students to navigate and fulfill the social and communicative purposes of writing within specific communities.

Unfortunately, students may rarely experience these facilitators of writing development, and many students lack sufficient writing skills, struggling to compose well-organized, coherent, elaborated texts written using a variety of sentence

structures, appropriate word choice, and solid command of conventions (spelling, grammar, punctuation, capitalization) [3]. This is troubling because writing well is necessary for success in K–12 and post-secondary settings. A common reason cited by educators as a barrier to intensifying writing instruction and increasing practice and feedback is the time required to read, evaluate, and provide feedback on student writing. Indeed, evaluating writing is time-consuming, and cognitively and emotionally taxing [4]. Absent opportunities to experience regular practice guided by effective feedback, students are unlikely to develop proficient writing skills.

Automated writing evaluation systems (AWE hereafter) – also referred to as Automated Essay Evaluation (AEE)[5] – leverage advances in artificial intelligence (AI), natural language processing (NLP), and machine learning to provide immediate automated scoring and feedback in the form of suggestions for improvement when revising, although some systems provide only feedback without scoring [6]. AWE systems can assess multiple writing features and do so with greater reliability and efficiency than humans, particularly linguistic, structural, and syntactic features [7]. AWE systems have been used productively in L1, L2, and EFL contexts to support language instruction, as they offer feedback and assessment on linguistic and structural elements crucial for learners across varied language backgrounds [8, 9]. As such, these tools have been proposed to help expedite the practice–feedback cycle necessary for writing development, thereby supporting practitioners in giving students high quality instruction and frequent practice and feedback [9-12].

AWE systems may also include electronic graphic organizers to support planning, peer review functionality to support revision, spelling and grammar feedback to support editing, and embedded skill-building lessons or game-based strategy instruction to support growth in writing skills [13, 14]. Most AWE systems provide teachers with data displays to support classroom assessment functions, such as identifying struggling writers, benchmarking, instructional planning, and progress monitoring [15, 16]. Thus, AWE is intended to support formative assessment [9, 17] and, in doing so, assist educators in improving students' writing skills [10, 12].

Parallel to its numerous perceived benefits, concerns have been raised about AWE's role in L1, L2, and EFL writing instruction. For example, AWE is unable to read and understand text in the same way a human does [18]. Also, although AWE evaluates higher-level aspects of writing like development of ideas, presence of key genre elements, and style, it is regarded as being superior at evaluating lower-level aspects of writing ability (e.g., language use, syntax, cohesion) [19].

Despite the nuanced perception of the applicability of AWE in writing instruction, its appeal and adoption continue to grow [20, 21], perhaps underscoring just how persistent a barrier evaluating student writing is for educators. Thus, it is critical that researchers, interventionists, and educators receive guidance on optimal uses of AWE because *adoption* of AWE will not in and of itself guarantee instructional benefits to teachers or students [9]. Those benefits are dependent upon its *effective implementation* within intervention and instruction.

Effective AWE implementation involves both educators and students. To see positive changes in writing performance associated with AWE, educators must assign multiple tasks using the AWE system while simultaneously holding students accountable for completing them. Indeed, Li [22] showed that the different approaches teachers' took to integrating AWE in their ESL classrooms produced very different patterns of AWE usage and effects. Ideally, this effective implementation of AWE systems will result in a favorable division of labor, with each entity (teacher and AWE) working in harmony to leverage each other's strengths and mitigate each other's weaknesses [23]. But such an effective implementation is not always achieved.

Therefore, in this article, we delve into prior research on AWE's affordances and shortcomings for supporting evidence-based writing instruction practices. We present four recommendations for effectively implementing AWE within intervention and instruction. These recommendations are intended to guide researchers in developing effective interventions and to assist practitioners who have adopted, or are considering adopting, AWE in their schools and classrooms.

2. Affordances and Shortcomings of AWE for Writing Instruction

Previous research on AWE for L1, L2, and EFL writing instruction has highlighted multiple benefits for students' writing outcomes with average effect sizes ranging from 0.38 to 0.98 [24-27]. For instance, the combination of teacher-led instruction with AWE has been shown to support improvements in students' writing motivation, writing self-efficacy, revising quality, and overall writing performance [9, 28, 29]. Also, research with L1, L2, and EFL writers across several age ranges indicates that AWE systems support reductions in the number of errors in students' writing [30], as well as improvements in the use of text evidence in essays [31], and more effective scientific argumentation [32].

However, the perceived benefits of AWE come with caveats. For example, research has shown that AWE systems tend to have less precision and recall than humans when detecting writing errors, which may contribute to limited AWE feedback uptake by students [30, 31, 33]. Moreover, Perelman [18] criticized automated scoring systems for being easily fooled to provide high ratings to an essay that uses sophisticated vocabulary and complex sentence structures despite the text being completely incoherent. Others have raised concerns that prolonged exposure to AWE will lead students to develop formulaic, or otherwise problematic, writing habits and to misunderstand the purpose of writing, which is social in nature [34].

Recognition of these limitations has generated concerns about the interpretations and uses of AWE scores (i.e., validity of AWE) [35], particularly in the elementary and middle grades when students are developing their understanding of what makes for "good" writing. Nevertheless, educators tend to regard AWE systems as usable and effective, and a proliferation of evidence indicates their overall positive contribution as a tool aiding in students' writing development [9].

Usability studies indicate that AWE may increase the amount of writing practice and feedback opportunities students receive [28]. Findings suggest that educators can manage their time more effectively and provide more focused feedback when using AWE [23] and reduce the amount of lower-level feedback they provide [30]. Moreover, educators indicate that AWE

facilitates classroom management: elementary students are kept productively engaged by AWE feedback, enabling the teacher to effectively conduct a 1:1 conference without interruption [36].

Despite teachers' general appreciation for AWE as a complementary tool for instruction, positive effects are not always salient in intervention studies. One reason for this is the aforementioned distinction between AWE's *adoption* and its *effective implementation*. If AWE is adopted but not implemented effectively within intervention and instruction, its potential to support teaching and learning is limited. Indeed, there is research to suggest that AWE may be underutilized in practice, with students practicing and revising to a limited degree [17]. Pressures to keep pace with curricula that de-prioritize writing or lack of coordinated administrative support [37] may partially explain under-utilization.

This scenario underscores the need to provide researchers, interventionists, and teachers with recommendations for planning intervention and instruction with AWE. In that spirit, we present the following four research-based recommendations for integrating AWE with evidence-based instructional practices: (a) clarify AWE's evaluation criteria, (b) teach writing strategies, (c) provide feedback, and (d) set goals and monitor progress.

3. Recommendations for Teaching Writing with AWE

3.1. Recommendation 1: Clarify evaluation criteria

According to MacArthur et al. [38], evaluation criteria comprise the qualities of effective writing that students should aspire to produce and against which their writing will be evaluated. When students have a strong understanding of evaluation criteria, it helps them more effectively revise and improve their writing—indeed the cognitive process of evaluation is central to the process of revision [39]. Moreover, clarifying evaluation criteria is the foundation of classroom formative assessment that later involves eliciting samples of student writing, evaluating those samples against the evaluation criteria, providing students with feedback to move them forward, engaging students as peer supports, and helping students take ownership of their learning [40]. Knowledge of evaluation criteria also provides students access to disciplinary writing communities [1].

Each AWE system uses specific criteria or rubrics to guide the scoring and feedback provided to students based on the genre and task type, and even the specific writing prompt. Ideally, repeated exposure to AWE's immediate scoring and feedback enables students to grasp what aspects are being assessed. As students apply these evaluation criteria in their revisions, they gain insights into qualities of effective writing, which they can then use when authoring future papers [9]. However, implicit learning alone may not be sufficient for all students; many may benefit from additional explicit instruction to help internalize the criteria, enabling them to more effectively interpret and apply AWE feedback, which some students may find confusing or unclear [41].

3.1.1. Actively engage students with the AWE evaluation criteria.

Educators utilizing AWE should ensure that students have a basic understanding of how to interpret the evaluation criteria and feedback comments provided by AWE. MacArthur et al. [38] identified an effective instructional process that can be used to teach AWE's evaluation criteria. The process begins with teacher-led discussion of what each of those criteria mean (e.g., "What does 'development of ideas' mean and look like in the context of argumentative writing?") along with introducing students to a rubric that elucidates the continuum of quality for those criteria. Subsequently, teachers engage in think-aloud modeling to read and evaluate strong and weak compositions while using a rubric. Following this process, students engage in collaborative, guided practice to evaluate a weak composition with teacher support. Finally, students evaluate a weak composition in small groups, pairs, or independently. This model of instruction involves a gradual release of responsibility to ensure students understand the meaning of those evaluation criteria and know how to apply them.

Some AWE systems include sample student essays that have been evaluated by the system. These sample essays would be ideal to use during the instructional process because not only would students come to better understand the evaluation criteria and how to apply them, but they would learn more about how the AWE system applies those criteria. If an AWE system does not have such essays, teachers should consider selecting strong and weak examples of student writing and submitting them to the AWE system for scoring and feedback. Comparing teacher and student evaluations of those texts to the AWE evaluation will not only reinforce knowledge of the evaluation criteria themselves, but it will also help students understand how the AWE system "thinks" about student writing. This type of evaluative triangulation is necessary because the exact linguistic features, and weighting of those features, used by the AWE system to generate its scoring and feedback are hidden.

3.2. Recommendation 2: Integrate AWE and strategy instruction

A meta-analysis of prior meta-analyses of writing instruction [2] revealed that teaching students writing strategies (i.e., a systematic procedure for completing a cognitive task) is one of the most powerful instructional practices for students. That same study also showed that strategy instruction is particularly effective when using the *self-regulated strategy development* model (SRSD hereafter). SRSD supports not only cognitive and metacognitive aspects of composing but affective and self-regulatory aspects, as well [42]. When students are self-regulated they work independently and self-monitor and self-evaluate as they work to achieve their goals. Graham and Harris [2] reported that those who received strategy instruction *and* SRSD produce better writing products when compared to their peers who received different forms of instruction.

AWE and SRSD complement each other because they both focus on applying the writing process to compose in different genres. Most AWE systems are designed to support the writing process—the iterative cognitive, metacognitive, and behavioral process of planning, translating and transcribing ideas into text, and evaluating, reviewing, and revising one's writing

to achieve a communicative purpose within a specific writing community [1, 39]. For instance, AWE systems often include graphic organizers to facilitate planning and always include automated feedback to scaffold productive reviewing and revising, helping students notice errors, clarify the nature of those errors, and identify effective repair strategies [43].

When SRSD is combined with an instructional feedback intervention, there are strong effects [44]. Given AWE's unique ability to provide immediate, nuanced feedback, integrating AWE with strategy instruction and SRSD shows great promise and has been shown to be effective [28]. Unlike traditional strategy instruction, where feedback may be delayed, AWE systems deliver immediate, nuanced, and iterative feedback that enables students to self-monitor and refine their writing continuously, an approach that aligns well with SRSD and enhances students' ability to apply strategies independently. Additionally, AWE systems offer consistent and scalable support for writing strategies across genres, alleviating the challenges of teacher-led feedback cycles in larger classes or time-limited settings and allowing for more structured, frequent practice.

However, just because AWE provides support for the writing process does not ensure that students will effectively utilize those supports to compose. Therefore, to capitalize on the strength of strategy instruction for writing, and to ensure students use AWE effectively, researchers, interventionists, and teachers should explicitly teach strategies for planning and drafting, revising, and editing using AWE. As illustrated by Palermo and Thomson [28] and in novel applications by Wijekumar et al. [45], by explicitly teaching composing strategies within AWE, educators can leverage AWE's scalability to ensure that all students, regardless of class size or instructor availability, receive structured, strategy-focused feedback—a challenge often faced in traditional settings.

How might this technological–pedagogical integration be achieved? First, educators can start by developing and activating students' background knowledge through thorough discussions of key elements of the genre and reviewing grade-level essays of varying quality together with students in the AWE system. Second, once students understand the writing strategy, they commit to using it. This can be achieved by motivating students to draft their essays using genre-specific elements, and to set clear writing goals. In the third stage, educators should model the strategy while thinking aloud, modeling how to select a suitable graphic organizer with the AWE system, fill out the electronic graphic organizer, and use that organizer to draft their essay.

Fourth, transferring responsibility to students involves ensuring that they remember the steps of the strategy. Quizzes or asking students to model the strategy to peers in small groups provides agency and tests students' assimilation of the strategy [2]. Once the strategy is memorized, the fifth stage is to support the strategy by providing plenty of opportunities for students to practice using the strategy while receiving feedback.

The sixth and final stage focuses on fostering the independent use of the strategy, alongside self-monitoring and self-evaluating their writing. Interventionists and teachers can support this by conducting individual writing conferences to supplement the AWE feedback—writing conferences are typically face-to-face interactions between teachers and students focusing on discussing the students' writing and providing individualized feedback [46]. When conferencing is conducted in an AWE-supported classroom, teachers should leverage AWE's feedback and data reporting functions to help students reflect on their progress and growth as writers.

To maximize the impact of integrating AWE and SRSD, developers should consider explicitly designing AWE systems to align with the SRSD instructional method. By embedding features that support SRSD's structured stages—from background knowledge activation to independent practice—AWE can complement traditional instruction while also extending SRSD's benefits in unique ways. For example, AWE's immediate feedback can reinforce students' understanding of genre-specific criteria at each stage, while data dashboards can help students self-monitor and self-evaluate across tasks. Indeed, researchers have begun developing such systems and have shown positive effects on students' writing outcomes [45]. Aligning AWE tools with SRSD not only leverages the strengths of both approaches but also provides educators with a practical and efficient way to deliver high-quality writing instruction.

3.3. Recommendation 3: Provide effective teacher feedback to supplement AWE feedback

Positive effect sizes on writing quality have been recorded by providing students with timely and effective feedback from adults, peers, oneself, or computers [2, 8]. Effective feedback helps students understand where they are and where they need to go [40]. Feedback also is an important source of information from which students derive perceptions of their self-efficacy [47], underscoring that the effects of feedback are not restricted to *the writing* but also influence *the writer*. Unfortunately, interventionists and educators may struggle to provide effective feedback because they lack resources, such as time and access to sufficiently nuanced, efficient, and reliable writing assessments. Educators also struggle to prioritize feedback on lower- and higher-level writing concerns, often emphasizing the former, despite that type of feedback being less effective for improving writing quality [48]. Perhaps for this reason, findings from studies of feedback in L1, L2, and EFL contexts, although generally positive [8], have been mixed [49].

3.3.1. Leverage the division of labor

AWE's efficiency and data reporting functions afford educators a means of providing more effective feedback via the creation of a supportive division of labor [36]. AWE systems are particularly adept at addressing low-level writing concerns [19] and can help students address such errors. However, AWE systems are not as adept at providing feedback on higher-level writing concerns, and many AWE systems do not evaluate the *meaning* of the writing, just the *quality* of the writing itself. This makes for a natural division of labor: AWE assists educators in providing students with feedback on lower-level writing concerns, so

that educators can provide proportionately more feedback on higher-level writing concerns and less feedback on lower-level skills [30].

3.3.2. *Use a mix of manners*

Moreover, the division of labor created when using AWE [30, 36] allows interventionists and teachers to focus more carefully on *how* they provide feedback. Teachers should consider providing feedback in different *manners* using a mix of *directives*, *queries*, *informatives*, and *praise* to maximize students' engagement with, and uptake of, teacher feedback [50]—see Table 1 for definitions and examples of different manners of providing feedback. AWE feedback also tends to target the task rather than the writing process or the writer's self-regulation. Thus, educators should consider supplementing AWE feedback with feedback that focuses on students' writing process and self-regulatory behaviors.

Table 1
Different manners of providing feedback

Feedback Manner	Definition	Benefits and Drawbacks	Examples
Directives	Teachers give explicit instructions to students on what to fix.	Directives are specific and easy to implement, but a student may become more reliant on the teacher.	"Put 'Hello Grace!' in quotation marks." "Provide a conclusion that restates your thesis."
Queries	Teachers ask students questions about the piece of writing.	Queries allow students to come to their own understandings, but this indirect feedback method may not be explicit enough to assist some writers.	"Do you think this was the best solution to the problem?" "What do you mean by this sentence?"
Informatives	Teachers make general comments about revisions without explicitly stating where the changes should occur.	Informatives remind students to look for certain things in their own writing. However, much like queries, these types of comments may not be sufficiently explicit.	"Remember that facts need to be supported by evidence that is cited from the text." "Good description does not tell the audience what is happening; it shows them."
Praise	Teachers tell students what they did well or what they improved.	Specific praise is useful for providing encouragement and motivation and helping to create a trusting environment when conferencing. However, praise does not inform students of what they need to do to move forward.	"You did a great job at using imagery to describe the setting!" "Your introduction has greatly improved since your first draft."

3.3.3. Focus on conferencing

Finally, educators should take advantage of the time- and labor-saving affordances, and classroom management benefits of AWE [36] to focus on building positive relationships with students during the writing conference. Individual conferences provide an excellent opportunity to encourage reluctant writers, shape positive writing mindsets, and build students' confidence that they can write, revise, and improve as writers. Furthermore, implementing peer review processes and collaborative writing can help develop a supportive writing community [2, 51].

3.3.4. Assign peer reviews

Peer feedback is an effective formative assessment practice [8] that benefits both the writer and the reviewer [52]: writers receive additional feedback and reviewers experience practice identifying writing problems and providing suggestions for improvement. Importantly, research suggests that both lower- and higher-ability writers benefit from peer feedback [53]. Many AWE systems support peer review, sometimes allowing anonymous peer review. Teachers should leverage this affordance to implement peer review in their classrooms. Doing so will afford students another source of feedback to supplement that of the AWE system [54], a combination shown to be effective for supporting students' writing development, including English learners [55].

However, assigning peer review within the AWE system will not ensure successful peer review. Students must be prepared to conduct peer review, a preparation that begins with instruction provided about evaluation criteria (Recommendation 1). Next, students should be taught how to apply evaluation criteria to evaluate each others' writing, identify areas of strength and improvement, and provide helpful feedback, first with modeling, then guided practice, and then collaborative and independent practice [38, 52]. When appropriately prepared to conduct peer review, students can provide effective feedback to supplement AWE feedback.

3.4. Recommendation 4: Use AWE to support goal setting and progress monitoring

Teaching students to set writing goals has a positive effect on students' overall writing quality [2]. For example, Ferretti et al. [56] demonstrated that presenting students with specific goals related to the inclusion of persuasive elements leads to better persuasive writing performance. AWE can support interventionists and teachers implementing goal setting with their students because AWE provides a means of monitoring progress via its scoring capabilities. Educators should encourage students using AWE to set and monitor progress related to three types of goals: performance goals, process goals, and product goals. Performance goals refer to the grade or score that the student aims to achieve on a given writing task. Goals that focus on how the writer goes about composing their text (e.g., planning, drafting, revising, and editing) are referred to as *process goals*, whereas goals that focus on the inclusion and development of specific writing features (e.g., text structure, sentence structure, word choice) are referred to as *product goals*. All three types of goals help students improve their writing and the quality of their revisions. Importantly, through goal setting teachers can help students develop positive, productive, and adaptive attributions and mindsets and self-efficacy [57].

When using AWE to set performance goals to stimulate students' motivation, students should be taught how to review their past performance collected and displayed within AWE electronic portfolios. Then, students should be taught how to set a reasonable performance goal for a subsequent writing task. Students can then identify specific relevant process and product goals that will serve as the concrete steps they will take to achieve their overall performance goal. Table 2 presents example performance, process, and product goals across key phases of the writing process [1, 39]. For instance, if a student's performance goal is to improve their performance by three points as measured by the AWE scoring system, their process goal(s) (e.g., "I will create a more detailed plan for my writing) and their product goal(s) (e.g., "I will use stronger word choice to better convey my opinion) will serve to direct students' activity in productive ways to increase the likelihood that they will achieve their performance goal. As students work to implement their process and product goals in pursuit of their performance goal, they can leverage AWE's feedback and scoring to support progress monitoring. A unique feature of AWE for progress monitoring is its ability to provide immediate and 100% consistent scoring, unlike human raters. This reliability allows for true measurement of growth on a consistent scale, making it easier for both students and educators to track incremental progress over time.

Table 2
Example performance, process, and product goals for students using AWE

Phase of the Writing Process	Example Performance Goals *Goals focusing on the grade or score the student aims to achieve*	Example Process Goals *Goals focusing on how the writer composes their text*	Example Product Goals *Goals focusing on the inclusion and development of specific writing features*
Planning and Drafting	• I scored a 15 on the first draft of my last persuasive writing assignment, so my goal this time is to score a 17 or above on my first draft.	• I will complete my full graphic organizer before I begin drafting. • I will double check my writing against my plan and make sure I have all the elements.	• My draft will include a clear opinion, three reasons, evidence for each reason, and a conclusion before I submit my draft persuasive writing for scoring.
Revising	• I will revise and improve my writing to get a score of 20 or more. • On my last assignment, I improved my first draft by three points when I revised. This time I will try to improve my first draft by four or more points.	• I will apply the feedback I receive from my teacher and AWE to improve my writing. • I will double-check my writing to ensure I have all the required genre elements and that each is well-developed	• My writing will include connections and transition words to better organize my ideas. • My writing will have a variety of sentence types.
Editing	• I will improve my sentence fluency score by 1 or more points when I revise. • I will improve my "Conventions" score by one or more points when I revise.	• I will read my text out loud to identify areas that sound awkward or unclear. • I will review each grammar and spelling suggestion I receive from the AWE system.	• All of my sentences will start with a capital letter and end with correct punctuation. • My writing will include correctly punctuated dialogue and quotations.

As students monitor progress towards the goals they set, AWE can play a key role in helping them connect their effort, actions, and outcomes. This is an important connection, as it helps to instill in students a growth versus fixed mindset approach to writing development—such implicit theories of writing development are important components of writing motivation and development [1, 58]. Interventionists and teachers can guide students to use AWE's scoring breakdowns and data displays to visualize how specific actions impact their results. For instance, during a writing conference, a teacher can help students see that the depth and breadth of their planning directly relate to the quality scores on their first draft—assignments with limited planning often receive lower initial scores compared to those with more developed plans [59]—and the inclusion of specific genre features improve their organization and idea development scores as rated by the AWE system. Without AWE's immediate, data-driven insights, it would be difficult, if not impossible, for teachers to consistently track and demonstrate these connections across large numbers of students. By referencing the consistent scoring data provided by AWE, teachers can reinforce this relationship, helping students recognize that sustained effort is reflected in measurable improvements across drafts.

4. Final Suggestions

Based on a review of prior research on AWE and evidence-based writing instruction practices, we provided four research-based recommendations for effectively implementing AWE within writing intervention and instruction. These recommendations leverage AWE's affordances for supporting L1, L2, and EFL writing development while mitigating its weaknesses and limitations. Each recommendation may be effective on its own, but ideally, all four recommendations will be implemented within the context of classroom writing routines or a writing intervention to create a powerful cycle of formative feedback and conferencing accelerated by the capabilities and affordances of AWE. In this way, AWE would be linked to formative assessment practices [40] that promote students' self-regulation and writing development [9].

After ensuring students understand AWE's underlying evaluation criteria (Recommendation 1), interventionists and educators should begin by integrating AWE with writing strategies, such as via SRSD (Recommendation 2). Then, they should use AWE to help provide more effective feedback (Recommendation 3). While conferencing, interventionists and teachers should focus specifically on providing higher-level feedback in varied manners (Recommendation 3) and should leverage AWE's feedback, consistent scoring, and data visualizations to help students set and monitor performance, product, and process goals (Recommendation 4).

However, these recommendations may be implemented in a different order. For instance, SRSD is a set of complex instructional practices that may require professional development to implement effectively [60]. If educators feel that strategy instruction, and specifically SRSD, seem overwhelming, starting off slower and building up to integrating AWE with strategy instruction in the writing process more broadly is recommended. In that case, educators could use AWE to provide more effective feedback (Recommendation 3) and support goal setting and progress monitoring (Recommendation 4). Eventually, educators can integrate AWE with more intensive instructional practices like strategy instruction and SRSD. Regular assessment and reflection on AWE integration can help educators refine their approach based on student progress and feedback. By continuously evaluating the impact of AWE-supported instruction, educators can identify which of these recommendations are most effective for their specific student populations and adjust their instructional practices accordingly.

5. Conclusion

AWE is an increasingly prevalent form of AI-supported educational technology. AWE's affordances may help educators overcome persistent barriers to intensifying writing instruction and increasing the amount of writing practice and feedback students experience. However, AWE is not a panacea; it has limitations that require mitigation by way of thoughtful implementation and integration within intervention and instruction. It is our hope that the recommendations provided in this article will help interventionists and teachers more effectively implement AWE with students to improve writing outcomes. Ultimately, it is our hope that as a result of such effective implementation, students will take ownership of their writing, becoming self-regulated writers who independently set goals, monitor progress, and refine their writing to achieve a communicative purpose within a writing community.

Funding

The research leading to the recommendations presented in this article was supported in part by Grant R305H170046 from the Institute of Education Sciences, U.S. Department of Education, to the University of Delaware. The opinions expressed in this paper are those of the authors and do not represent the views of the Institute, or the U.S. Department of Education, and no official endorsement by these agencies should be inferred.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Joshua Wilson: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Tania Cruz Cordero:** Writing – review & editing. **Andrew Potter:** Writing – review & editing. **Matthew Myers:** Writing – review & editing. **Charles A. MacArthur:** Writing – review & editing, Funding acquisition. **Gaysha Beard:** Investigation, Funding acquisition. **Emily A. Fudge:** Investigation, Writing – original draft. **Alexandria Raiche:** Investigation, Writing – original draft. **Cristina Ahrendt:** Investigation, Writing – original draft.

References

[1] Graham, S. (2023). Writer(s)-within-community model of writing as a lens for studying the teaching of writing. In R. Horowitz (Ed.), *The Routledge International Handbook of Research on Writing* (pp. 337–350). Routledge.

[2] Graham, S., & Harris, K. R. (2017). Evidence-based writing practices: A meta-analysis of existing meta-analyses. In R. Fidalgo, K. R. Harris, & M. Braaksma (Eds.), *Design principles for teaching effective writing: Theoretical and empirical grounded principles* (pp. 13-37). Brill. https://doi.org/10.1163/9789004270480_003

[3] National Center for Education Statistics. (2012). *The Nation's Report Card: Writing 2011* (NCES 2012–470). https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf

[4] Baker, N. L. (2014). "Get it off my stack": Teachers' tools for grading papers. *Assessing Writing, 19*, 36–50. https://doi.org/10.1016/j.asw.2013.11.005

[5] Shermis, M. D., & Wilson, J. (Eds.). (2024). *The Routledge international handbook of automated essay evaluation*. USA: Taylor & Francis.

[6] Cotos, E. (2023). Automated feedback on writing. In O. Kruse, C. Rapp, C. M. Anson, K. Benetos, E. Cotos, A. Devitt, & A. Shibani (Eds.), *Digital writing technologies in higher education: Theory, research, and practice* (pp. 347–364). Springer, Cham. https://doi.org/10.1007/978-3-031-36033-6_22

[7] Deane, P. (2024). Writing trait analysis. In M. D. Shermis & J. Wilson (Eds.), *The Routledge International Handbook of Automated Essay Evaluation* (pp. 278–302). Routledge.

[8] Scherer, S., Graham, S., & Busse, V. (2024). How effective is feedback for L1, L2, and FL learners' writing? A meta-analysis. *Learning and Instruction, 93*, 101961. https://doi.org/10.1016/j.learninstruc.2024.101961

[9] Wilson, J., & MacArthur, C. A. (2024). Exploring the role of automated writing evaluation as a formative assessment tool supporting self-regulated learning in writing. In M. D. Shermis & J. Wilson (Eds.), *The Routledge International Handbook of Automated Essay Evaluation* (pp. 197–220). Routledge.

[10] Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning, 37*(1-2), 179-221. https://doi.org/10.1080/09588221.2022.2033787

[11] Hockly, N. (2019). Automated writing evaluation. *ELT Journal, 73*(1), 82–88. https://doi.org/10.1093/elt/ccy044

[12] Shi, Z., Liu, F., Lai, C., & Jin, T. (2022). Enhancing the use of evidence in argumentative writing through collaborative processing of content-based automated writing evaluation feedback. *Language Learning & Technology*, *26*(2), 106–128. https://doi.org/10125/73481

[13] Butterfuss, R., Roscoe, R. D., Allen, L. K., McCarthy, K. S., & McNamara, D. S. (2022). Strategy uptake in writing pal: Adaptive feedback and instruction. *Journal of Educational Computing Research, 60*(3), 696–721. https://doi.org/10.1177/07356331211045304

[14] Fang, Y., Roscoe, R. D., & McNamara, D. S. (2023). Artificial intelligence-based assessment in education. In B. du Boulay, A. Mitrovic, & K. Yacef (Eds.), *Handbook of artificial intelligence in education* (pp. 485–504). Edward Elgar Publishing. https://doi.org/10.4337/9781800375413.00033

[15] Keller-Margulis, M. A., Mercer, S. H., & Matta, M. (2021). Validity of automated text evaluation tools for written-expression curriculum-based measurement: A comparison study. *Reading and Writing, 34*(10), 2461-2480. https://doi.org/10.1007/s11145-021-10153-6

[16] Mercer, S. H., Cannon, J. E., Squires, B., Guo, Y., & Pinco, E. (2021). Accuracy of automated written expression curriculum-based measurement scoring. *Canadian Journal of School Psychology*, *36*(4), 304–317. https://doi.org/10.1177/0829573520987753

[17] Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment, 8*(6), 1-44.

[18] Perelman, L. (2020). The BABEL generator and e-Rater: 21st century writing constructs and automated essay scoring (AES). *Journal of Writing Assessment, 13*(1). https://journalofwritingassessment.org/article.php?article=145

[19] Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18*(1), 7–24. https://doi.org/10.1016/j.asw.2012.10.002

[20] Barrot, J. S. (2024). Trends in automated writing evaluation systems research for teaching, learning, and assessment: A bibliometric analysis. *Education and Information Technologies, 29*, 7155–7199. https://doi.org/10.1007/s10639-023-12083-y
[21] Shi, H., & Aryadoust, V. (2023). A systematic review of automated writing evaluation systems. *Education and Information Technologies, 28*, 771–795. https://doi.org/10.1007/s10639-022-11200-7
[22] Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System, 99*, 102505. https://doi.org/10.1016/j.system.2021.102505
[23] Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research, 42*(2), 173–196. https://doi.org/10.2190/EC.42.2.c
[24] Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence.* 6:1162454. https://doi.org/10.3389/frai.2023.1162454
[25] Li, R. (2023). Still a fallible tool? Revisiting effects of automated writing evaluation from activity theory perspective. *British Journal of Educational Technology, 54*(3), 773–789. https://doi.org/10.1111/bjet.13294
[26] Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2022). The effectiveness of automated writing evaluation in EFL/ESL writing: A three level meta-analysis. *Interactive Learning Environments, 32*(2), 727–744. https://doi.org/10.1080/10494820.2022.2096642
[27] Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. L. (2022). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning, 38*(2), 599–620. https://doi.org/10.1111/jcal.12635
[28] Palermo, C., & Thomson, M. M. (2018). Teacher implementation of self-regulated strategy development with an automated writing evaluation system: Effects on the argumentative writing performance of middle school students. *Contemporary Educational Psychology, 54*, 255–270. https://doi.org/10.1016/j.cedpsych.2018.07.002
[29] Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51–65. https://doi.org/10.1016/j.asw.2013.11.007
[30] Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning, 35*(4), 605–634. https://doi.org/10.1080/09588221.2020.1743323
[31] Wang, E. L., Matsumura, L. C., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A., & Quintana, R. (2020). *eRevis(ing)*: Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing, 44*, 100449. https://doi.org/10.1016/j.asw.2020.100449
[32] 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, 143*, 103668. https://doi.org/10.1016/j.compedu.2019.103668
[33] Bai, L., & Hu, G. (2016). In the face of fallible AWE feedback: How do students respond? *Educational Psychology, 37*(1), 67–81. https://doi.org/10.1080/01443410.2016.1223275
[34] Conference on College Composition and Communication. (2014). *Writing assessment: A position statement.* Retrieved from https://cccc.ncte.org/cccc/resources/positions/writingassessment.
[35] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing.* USA*:* American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
[36] Wilson, J., Ahrendt, C., Fudge, E., Raiche, A., Beard, G., & MacArthur, C. A. (2021). Elementary teachers' perceptions of automated feedback and automated scoring: Transforming the teaching and learning of writing using automated writing evaluation. *Computers & Education, 168*, 104208. https://doi.org/10.1016/j.compedu.2021.104208
[37] Mayfield, E., & Butler, S. (2018). Districtwide implementations outperform isolated use of automated feedback in high school writing. In *Proceedings of the International Conference of the Learning Sciences* (pp. 23-27). http://ceur-ws.org/Vol-2128/industrial4.pdf
[38] MacArthur, C. A., Philippakos, Z. A. T., May, H., & Compello, J. (2022). Strategy instruction with self-regulation in college developmental writing courses: Results from a randomized experiment. *Journal of Educational Psychology, 114*(4), 815–832. https://doi.org/10.1037/edu0000705
[39] Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 1–27). Lawrence Erlbaum Associates.
[40] Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5–31. https://doi.org/10.1007/s11092-008-9068-5
[41] Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing, 52*, 100816. https://doi.org/10.1016/j.jslw.2021.100816
[42] Harris, K. R., & McKeown, D. (2022). Overcoming barriers and paradigm wars: Powerful evidence-based writing instruction. *Theory Into Practice, 61*(4), 429–442. https://doi.org/10.1080/00405841.2022.2107334
[43] Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning, 36*(4), 584–607. https://doi.org/10.1080/09588221.2021.1936071

[44] Nunes, A., Cordeiro, C., Rocha, R., Limpo, T., & Castro, S. L. (2024). "Breath, plan, write, and evaluate": The effects of an SRSD intervention and instructional feedback on 4th graders' writing and motivation. *Frontiers in Education, 9*. https://doi.org/10.3389/feduc.2024.1305771

[45] Wijekumar, K., McKeown, D., Zhang, S., Lei, P.–W., Hruska, N., & Pirnay–Dummer, P. (2024). We Write automated scoring. In M. D. Shermis & J. Wilson (Eds.), *The Routledge International Handbook of Automated Essay Evaluation* (pp. 178–194). Routledge.

[46] Patthey–Chavez, G. G., & Ferris, D. R. (1997). Writing conferences and the weaving of multi-voiced texts in college composition. *Research in the Teaching of English, 31*(1), 51–90. https://www.jstor.org/stable/40171264

[47] Dujinhower, H., Prins, F. J., & Stokking, K. M. (2010). Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance. *Learning and Instruction, 22*(3), 171–184. https://doi.org/10.1016/j.learninstruc.2011.10.003

[48] Matsumura, L. C., Patthey-Chavez, G. G., Valdés, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower and higher-achieving urban schools. *The Elementary School Journal, 103*(1), 3–25. https://doi.org/10.1086/499713

[49] Biber, D., Nekrasova, T., & Horn, B. (2011). *The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. TOEFL iBT™ research report.* https://files.eric.ed.gov/fulltext/EJ1111175.pdf

[50] Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing, 51*, 100586. https://doi.org/10.1016/j.asw.2021.100586

[51] Friedrich, L. (2019). Setting up the writing classroom. In S. Graham, C. A. MacArthur, & M. Hebert (Eds.), *Best practices in writing instruction* (3rd ed., pp. 31-50). The Guilford Press.

[52] Philippakos, Z. A. (2017). Giving feedback: preparing students for peer review and self-evaluation. *The Reading Teacher, 71*(1), 1, 13–22. https://doi.org/10.1002/trtr.1568

[53] Patchan, M. M., & Schunn, C. D. (2016). Understanding the effects of receiving peer feedback for text revision: Relations between author and reviewer ability. *Journal of Writing Research, 8*(2), 227–265. https://doi.org/10.17239/jowr-2016.08.02.03

[54] Li, W. Y., Kau, K., & Shiung, Y. J. (2023). Pedagogic exploration into adapting automated writing evaluation and peer review integrated feedback into large-sized university writing classes. *SAGE Open, 13*(4). https://doi.org/10.1177/21582440231209087

[55] Liaqat, A., Munteanu, C., & Demmans Epp, C. (2021). Collaborating with mature English language learners to combine peer and automated feedback: A user-centered approach to designing writing support. *International Journal of Artificial Intelligence in Education, 31*, 638–679. https://doi.org/10.1007/s40593-020-00204-4

[56] Ferretti, R. P., Lewis, W. E., & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies. *Journal of Educational Psychology, 101*(3), 377–589. https://doi.org/10.1037/a0014702

[57] Chung, H. Q., Chen, V., & Olson, C. B. (2021). The impact of self-assessment, planning and goal setting, and reflection before and after revision on student self-efficacy and writing performance. *Reading and Writing, 34*, 1885–1913. https://doi.org/10.1007/s11145-021-10186-x

[58] Camacho, A., Alves, R. A., & Boscolo, P. (2021). Writing motivation in school: A systematic review of empirical research in the early twenty-first century. *Educational Psychology Review, 33*, 213–247. https://doi.org/10.1007/s10648-020-09530-4

[59] Cordeiro, C., Limpo, T., Olive, T., & Castro, S. L. (2020). Do executive functions contribute to writing quality in beginning writers? A longitudinal study with second graders. *Reading and Writing, 33*, 813–833. https://doi.org/10.1007/s11145-019-09963-6

[60] Harris, K. R., Camping, A., & McKeown, D. (2023). A review of research on professional development for multicomponent strategy-focused writing instruction: Knowledge gained and challenges remaining. In F. DeSmedt, R. Bouwer, T. Limpo, & S. Graham (Eds). *Conceptualizing, designing, implementing, and evaluating writing interventions* (pp. 101–136). Brill Publishing.

How to Cite

Wilson, J., Cruz Cordero, T., Potter, A., Myers, M., MacArthur, C. A., Beard, G., Fudge, E. A., Raiche, A., & Ahrendt, C. (2024). Recommendations for Integrating Automated Writing Evaluation with Evidence-Based Instructional Practice. *International Journal of Changes in Education*. https://doi.org/10.47852/bonviewIJCE42024011