

The Writing Analytics Tool: A Learning Engineering Approach to Designing AI-Supported Writing Instruction

McNamara, D. S., Potter, A., Serhan, Z., Singh, M., Patne, N., Arner, T., Roscoe, R. D., Allen, L. K. , & Crossley, S. A.

AI in Education

automated writing evaluation

learning engineering

Participatory Design

Writing Analytics

The Writing Analytics Toolkit (WAT) is an AI-driven platform designed to support writing instruction and research through transparent, theory-grounded writing analytics. Developed over six years through an IES-funded learning engineering initiative, WAT integrates natural language processing, machine learning, participatory design, and evidence-based instructional principles to serve students, teachers, and researchers within a single system. This paper describes WAT's architecture, analytics, and interfaces, and frames its development as a learning engineering case study. We document how stakeholder engagement, iterative design cycles, empirical validation, and scalability considerations shaped WAT's evolution. The work illustrates how learning engineering can guide the responsible design of AI-enabled educational tools that are pedagogically aligned, usable in authentic contexts, and extensible for research and innovation.

Introduction

Writing is a foundational academic and professional skill, yet it remains difficult to teach, assess, and study at scale (Graham, 2019). Large class sizes, limited instructional time, and the labor-intensive nature of writing feedback continue to constrain opportunities for deliberate practice and revision (e.g., Applebee & Langer, 2011). Automated writing evaluation (AWE) systems have emerged as a partial response, but many commercial AWE systems operate as opaque “black boxes,” offering scores without transparency or pedagogical grounding (McNamara & Potter, 2024; Wilson & MacArthur, 2024). In response to these challenges, the Writing Analytics Toolkit (WAT) was developed as an open, research-driven platform that provides actionable writing analytics to students, teachers, and researchers.

This paper has two goals. First, we describe WAT as it exists today, including its tools, analytics, and empirical foundations. Second, we frame WAT's development as a learning engineering process, illustrating how sustained stakeholder engagement, iterative design cycles, and evidence-centered analytic practices were used to align AI capabilities with instructional goals. In doing so, we position WAT as a concrete example of how learning engineering can guide the responsible development of AI-enabled educational technologies.

Learning Engineering as a Design Framework

Learning engineering is an interdisciplinary approach to designing and improving learning systems that integrates learning science, data-informed analytics, and human-centered engineering practices, emphasizing the translation of research insights into implemented, scalable learning solutions in authentic contexts (Baker et al., 2022; Goodell & Kolodner, 2023). Learning engineering is best understood as a process organized around defining a learning challenge and engaging in iterative cycles of creation, implementation, and investigation, in which learning solutions, analytics, and infrastructure are designed, deployed,

and examined using empirical evidence (Kessler et al., 2023). These cycles are iterative rather than linear, with evidence from each phase informing subsequent design decisions.

This process is particularly important for the development of AWE systems. Research findings on AWE suggest that the instructional value of these systems is contingent on meaningful integration into writing instruction and investment by writing instructors (Potter et al., 2025b; Wilson et al., 2025). Ongoing challenges in AWE include limited transparency, uneven instructional uptake, and tensions between automation and pedagogy. These challenges underscore the need for design approaches that foreground context and human decision-making (Shermis & Wilson, 2024). As AWE systems increasingly incorporate methods from artificial intelligence (AI), including large language models (LLMs), questions of human-centered design and evidence-informed implementation remain central (McNamara & Potter, 2024). These questions motivate the use of learning engineering as a framework for examining how writing technologies can be developed through iterative cycles of challenge definition, creation, implementation, and investigation to support effective, human-centered instruction at scale.

As a first step in applying the learning engineering process to WAT, we begin by defining the instructional and research challenges that motivated the development of the system and describing how the current versions of the system seek to address those challenges.

Defining the Challenge: Writing Instruction and Research at Scale

A central principle of the learning engineering process is the explicit definition of the learning challenge, which guides design decisions across subsequent cycles of creation, implementation, and investigation. In developing WAT, our research team identified two closely related challenges that persist in writing instruction and writing research: 1) supporting deliberate writing practice for students with meaningful feedback at scale, and 2) advancing the study of writing development through scalable, theory-informed analytic methods. As such, WAT was designed as a platform comprising two complementary systems: WAT Classroom (WAT-C) and WAT Researcher (WAT-R).

The current version of WAT-C is a web-based system that supports formative writing instruction for high school students, college students, and their instructors by integrating writing analytics into existing instructional workflows. Instructors use WAT-C to create writing tasks, select and configure analytics aligned with their instructional goals, and review student submissions across drafts. The instructor interface emphasizes control and flexibility, allowing teachers to determine which analytics are visible to students and how feedback is framed, thereby preserving instructors' roles as the primary evaluators of writing quality. Students interact with WAT-C through a dashboard that supports assignment management, review of analytic feedback, and submission of revised drafts. Across both interfaces, analytics are presented descriptively rather than evaluatively, with the goal of supporting interpretation, reflection, and revision rather than score optimization.

Whereas WAT-C is designed for instructional use in classrooms, WAT-R is intended for post hoc analyses of writing corpora. The current version of WAT-R is a natural language processing (NLP) tool that extracts linguistic features and component-level metrics derived from student writing. WAT-R is implemented as a standalone desktop application for macOS and Windows, designed to be usable without coding or technical setup. Researchers upload a corpora of student writing and use WAT-R to extract linguistic features and component-level metrics in a standardized format. These outputs are intended to support empirical analyses of writing development, instructional effects, and individual differences, as well as integration with statistical and machine learning workflows. WAT-R facilitates the development and testing of more advanced theoretical models of writing while promoting comparability across studies by reducing the time and technical expertise required for large-scale linguistic analysis.

In the sections that follow, we use the design and development of WAT-C as a case study to describe how learning engineering cycles of stakeholder engagement, analytic development, implementation, and investigation can be leveraged to develop human-centered technologies. Further, we discuss how these cycles continue to guide evaluation and refinement of the system's response to the persistent challenges in writing instruction and writing research.

Creation: Stakeholder Engagement through Participatory Design

Effective and sustainable classroom tools must align not only with theoretical models of learning but also with the practical, contextual realities of instruction. Instructional tools that are designed without sustained input from educators may fail to integrate into classroom workflows, may be misaligned with pedagogical goals, or introduce unintended burdens that limit adoption and impact. Therefore, considering stakeholders, particularly instructors, as active contributors during development is critical for ensuring usability, instructional relevance, and long-term sustainability. Participatory and co-design approaches support this alignment by foregrounding practitioner expertise, surfacing tacit instructional knowledge, and enabling design decisions that reflect authentic classroom constraints and opportunities.

Guided by these principles, the WAT-C development team purposely and systematically implemented a participatory research design (Wacnik et al., 2025) to develop WAT-C in collaboration with experienced writing instructors (see Li et al., 2022 for details). Secondary writing teachers were engaged as co-designers to ensure that early system decisions reflected instructional practice and classroom realities. In addition to teacher participants, the development team included researchers with expertise in psychology, applied linguistics, writing assessment, and learning analytics, as well as software engineers and UX designers. Several members of the research team also had prior experience designing automated writing evaluation systems, including the Writing Pal intelligent tutoring system (Roscoe et al., 2014).

Through this stakeholder-centered creation process, a central design requirement became clear: instructors did not want WAT-C to replicate or approximate predictive evaluation scores. Instead, they wanted analytics that could help them better understand patterns in students' writing and use that information to support instruction and revision. From a learning engineering perspective, this input prompted a design pivot away from formative scoring and toward descriptive writing analytics that characterize features of students' texts and support interpretation rather than judgment. These analytics were intended to function as sensemaking tools, enabling instructors to learn more about students' writing and enabling students to reflect on and revise their own work. Teachers further informed the design of system affordances emphasizing methods to communicate personalized feedback to their students. These stakeholder-driven decisions established the foundation for subsequent learning engineering cycles focused on analytic development, implementation, and empirical investigation.

Creation and Investigation: Iterative Analytics Development

Following the shift toward descriptive analytics, learning engineering cycles focused on developing analytic instrumentation that could characterize features of student writing at scale while remaining interpretable for instructional use. To accomplish this goal, WAT's analytic infrastructure integrates a set of established, open-source natural language processing tools from the [SALAT](#) (i.e., the Suite of Automatic Linguistic Analysis Tools, Crossley & Kyle, 2018). These tools, which have been validated and widely used in prior writing research, support the extraction of lexical, syntactic, and discourse-level features of student writing, including indices related to lexical sophistication and diversity, syntactic complexity, and local and global cohesion, as well as genre-relevant features for persuasive and source-based writing such as claims, citations, quotations, and indicators of source overlap (see Potter et al., 2025a).

Given the large number of linguistic features produced by this infrastructure, principal component analysis (PCA) modeling approaches were used to organize analytics into interpretable dimensions of writing that could be used in WAT-C to provide meaningful feedback to instructors and students. PCA was applied to hundreds of correlated features extracted from a corpus of persuasive essays, source-based essays, and summaries that were reduced into a smaller set of theoretically meaningful components, such as academic language use, elaboration, cohesion, and language variety (see Potter et al., under review). These components were validated against human judgments of writing quality and conceptualized as descriptive and facilitative representations of writing features rather than evaluative scores.

Subsequent studies refined and extended these analytics through iterative investigation. For source-based writing, a source integration construct was developed by modeling linguistic features related to citation practices, quotation, plagiarism

indicators, and semantic overlap with source texts, with evidence of strong alignment with human ratings and generalizability across prompts and datasets (Potter et al., under review). Related work developed and validated a global cohesion metric and extended its use to evaluate revisions generated by LLMs, illustrating how validated linguistic analytics can be applied to emerging instructional contexts while maintaining construct validity (Potter et al., in press b).

Implementation and Investigation: System Refinement and Scaling

Following the development of WAT's writing analytics, subsequent learning engineering cycles focused on extended investigation and system refinement to assess feasibility, usability, and instructional relevance prior to broader deployment. A second participatory design study in collaboration with writing instructors examined how WAT's analytics, workflows, and configuration options aligned with instructional needs (see Potter et al., in press a). Instructor feedback informed targeted refinements to metric presentation and customization features, as well as principled decisions about scope. Notably, instructors reported limited instructional use of summaries, reflecting shifts in writing practice in the context of generative AI tools. In combination with emerging evidence that summaries may offer weaker support for writing development relative to other genres (McNamara et al., 2024), these findings motivated the removal of summaries from the instructional interface in favor of deeper support for persuasive and source-based writing.

In parallel, system-level refinements addressed scalability and readiness for deployment. WAT's infrastructure was re-architected to improve computational efficiency and support future integration with learning management systems through secure cloud-based services and selective index computation. The cloud infrastructure was designed and tested to securely manage user data, support concurrent access by multiple users, and scale dynamically under increased load, ensuring reliability and data protection as adoption increases. These changes were designed to enable real-time student feedback while minimizing instructional and technical burden.

Building on these refinements, the next learning engineering cycle focuses on integrating generative AI to support students' interpretation and use of writing analytics during drafting and revision. Instructor feedback from the participatory design study indicated that students increasingly expect personalized, conversational support aligned directly with their own writing, and that instructors would be hesitant to adopt analytics tools that do not incorporate such affordances. In response, ongoing work explores the integration of a chatbot powered by a large language model to scaffold writing and revision using WAT-produced analytics, while simultaneously maintaining human-in-the-loop control over feedback and evaluation. Planned experimental studies will examine how students interact with this chatbot and evaluate its effects on revision quality, writing outcomes, and instructional use, extending WAT's learning engineering cycle through systematic investigation of this next refinement.

Conclusion

The Writing Analytics Toolkit case study illustrates how learning engineering can guide the development of AI-enabled educational tools over extended timelines. Across its development, learning engineering principles shaped decisions about analytic representation, system architecture, and refinement, resulting in a platform that prioritizes interpretability, instructional relevance, and usability alongside scalability.

As both writing instruction and educational research increasingly intersect with advances in artificial intelligence, WAT provides an example of how AI-supported tools can be designed to augment human judgment and practice opportunities in educational contexts. More broadly, this work underscores the value of learning engineering as a framework for ensuring that emerging educational technologies remain aligned with evidence-based instructional practices and user needs. Continued

investigation of WAT and its future iterations will further inform how learning-engineered tools can responsibly support teaching and learning.

Acknowledgments

This work was supported by the Institute of Education Sciences (R305A180261) as well as the Learning Engineering Institute at Arizona State University. We thank the National Writing Project teachers, instructors, students, and research team members who contributed to WAT's development.

References

Applebee, A. N., & Langer, J. A. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal*, 100(6), 14–27. <https://doi.org/10.58680/ej201116413>

Baker, R. S., Boser, U., & Snow, E. L. (2022). Learning engineering: A view on where the field is at, where it's going, and the research needed. *Technology, Mind, and Behavior*, 3, 1–23. <https://doi.org/10.1037/tmb0000058>

Crossley, S. A., & Kyle, K. (2018). Assessing writing with the tool for the automatic analysis of lexical sophistication (TAALES). *Assessing Writing*. 38 (1), 46-50.

Goodell, J., Kessler, A., & Schatz, S. (2023). Learning engineering at a glance. *Journal of Military Learning, Conference Edition*, 46–59.

Goodell, J., Kolodner, J. L., & Kessler, A. (2023). Learning engineering applies the learning sciences. In J. Goodell & J. L. Kolodner (Eds.), *Learning engineering toolkit: Evidence-based practices from the learning sciences, instructional design, and beyond*. Routledge. <https://doi.org/10.4324/9781003276579>

Graham, S. (2019). Changing how writing is taught. *Review of Research in Education*, 43(1), 277–303. <https://doi.org/10.3102/0091732X18821125>

Kessler, A., Craig, S. D., Goodell, J., Kurzweil, D., & Greenwald, S. W. (2023). Learning engineering is a process. In J. Goodell & J. L. Kolodner (Eds.), *Learning engineering toolkit: Evidence-based practices from the learning sciences, instructional design, and beyond*. Routledge. <https://doi.org/10.4324/9781003276579>

McNamara, D. S., & Potter, A. (2024). The two U's in the future of automated essay evaluation: Universal access and user-centered design. In *The Routledge international handbook of automated essay evaluation* (pp. 590–608). Routledge.

McNamara, D. S., Watanabe, M., Huynh, L., McCarthy, K. S., Allen, L. K., & Magliano, J. P. (2024). Summarizing versus rereading multiple documents. *Contemporary Educational Psychology*, 76, 102238. <https://doi.org/10.1016/j.cedpsych.2023.102238>

Potter, A. H., Crossley, S. A., Öncel, P., Allen, L. A., Xu, X., Balyan, R., Arner, T., Roscoe, R. D., & McNamara, D. S. (manuscript under review). From words to components: Mapping linguistic features in source-based writing.

Potter, A. H., Serhan, Z., Patne, N. A., Öncel, P., Ahmed, I., Arner, T., Islam, R., Allen, L. A., Roscoe, R. D., Crossley, S. A., & McNamara, D. S. (in press a). Human-AI Collaboration in Participatory Design: Refining the Writing Analytics Tool. *Journal of Educational Data Mining*.

Potter, A., Shortt, M., Goldshtain, M., & Roscoe, R. D. (2025a). Assessing academic language in tenth-grade essays using natural language processing. *Assessing Writing*, 64, 100921. <https://doi.org/10.1016/j.asw.2025.100921>

Potter, A., Tian, Y., Balyan, R., Watanabe, M., & McNamara, D. S. (in press b). Revising with generative AI: Supporting writing instruction through validated cohesion metrics. In Z. Philippakos (Ed.), *AI literacy: Foundations, practices, and innovations in education*. Routledge.

Potter, A., Wilson, J., Roscoe, R.D., Arner, T., & McNamara, D.S. (2025b). Computer-based writing instruction. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.). *Handbook of writing research* (3rd ed.). Guilford Press.

Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A., & McNamara, D. S. (2014). The Writing Pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, 34, 39-59. <https://doi.org/10.1016/j.compcom.2014.09.002>

Shermis, M. D., & Wilson, J. (2024). Introduction to automated essay evaluation. In *The Routledge international handbook of automated essay evaluation* (pp. 3–22). Routledge.

Thai, K. P., Craig, S. D., Goodell, J., Lis, J., Schoenherr, J. R., & Kolodner, J. L. (2023). Learning engineering is human-centered. In J. Goodell & J. L. Kolodner (Eds.), *Learning engineering toolkit: Evidence-based practices from the learning sciences, instructional design, and beyond*. Routledge. <https://doi.org/10.4324/9781003276579>

Wilson, J., & MacArthur, C. (2024). Exploring the role of automated writing evaluation as a formative assessment tool supporting self-regulated learning in writing. In *The Routledge international handbook of automated essay evaluation* (pp. 197–220). Routledge.

Wilson, J., Palermo, C., Myers, M. C., Cruz Cordero, T., Eacker, H., Potter, A., Coles, J., May, H., & Zhang, F. (2025). Impact of automated writing evaluation (AWE) on middle school students' writing outcomes. *American Educational Research Journal*. <https://doi.org/10.3102/00028312251388655>

