Automatic Identification and Evaluation of Revisions in Student Writing Using Large Language Models

Tian, Y., Patil, P. , Potter, A., Early, J., Graham, S., Christhilf, K., Ahandani, M. D., & McNamara, D.S.

Large Language Models   Revision Evaluation

Extended Abstract

Revision provides a window into writers' metacognitive awareness, goal-directed decision making, and linguistic development (Hayes, 2012). Understanding how students revise is therefore critical for both writing research and instruction. However, identifying and evaluating revisions remains challenging: manual annotation is labor-intensive and inconsistent, while existing automated approaches rely largely on surface metrics (e.g., deletion/insertion counts, edit distance) (Tian & Cushing, 2025). These limitations prevent researchers and educators from capturing the semantic and functional nature of revisions, ultimately constraining the usefulness of revision analytics for scalable, instructionally meaningful feedback. Advancing automated methods for identifying and evaluating revisions is essential for generating evidence-based feedback and supporting personalized writing development at scale.

This project introduces an automatic workflow that leverages large language models (LLMs) to identify, categorize, and evaluate revisions between students' draft and revised essays (See Appendix A for an overview). The workflow includes four stages: (1) detecting text changes between draft:revision pairs using rule-based approaches, (2) prompting an LLM to explain each change in natural language, (3) using an LLM to categorize revision types (e.g., typographic, grammatical, or content-level), (4) querying an LLM to evaluate each revision as Good, Neutral, or Bad and provide explanation. Two trained human annotators independently code revisions using a custom rubric (Appendix B). The annotations serve as ground truth for evaluating the accuracy and reliability of the automated workflow.

Initial analyses indicate that, with carefully designed prompts, state-of-the-art LLMs can reliably identify revision types and evaluate their quality using the textual context. We anticipate that our workflow will achieve performance comparable to human annotation and demonstrate strong potential for integration into writing research and instructional tools.

By harnessing LLMs for automated text revision analysis and evaluation, this work advances the development of evidence-based, scalable systems for writing analytics and personalized feedback. The workflow embodies the principles of learning engineering by uniting theory, data, and iterative design to model and support complex learning processes. Anticipated applications include formative feedback systems that highlight meaningful revisions, instructor dashboards that visualize students' writing growth, and adaptive learning environments that guide students toward higher-quality revisions.

Acknowledgments

References

Hayes, J. R. (2012). Modeling and remodeling writing. Written communication, 29(3), 369-388.

Tian, Y., & Cushing, S. T. (2025). Exploring the application of keystroke logging techniques to research in second language (L2) writing. Research Methods in Applied Linguistics, 4(1), 100179.

Appendix A

Figure A1

Overview of the automated workflow for revision identification and evaluation.
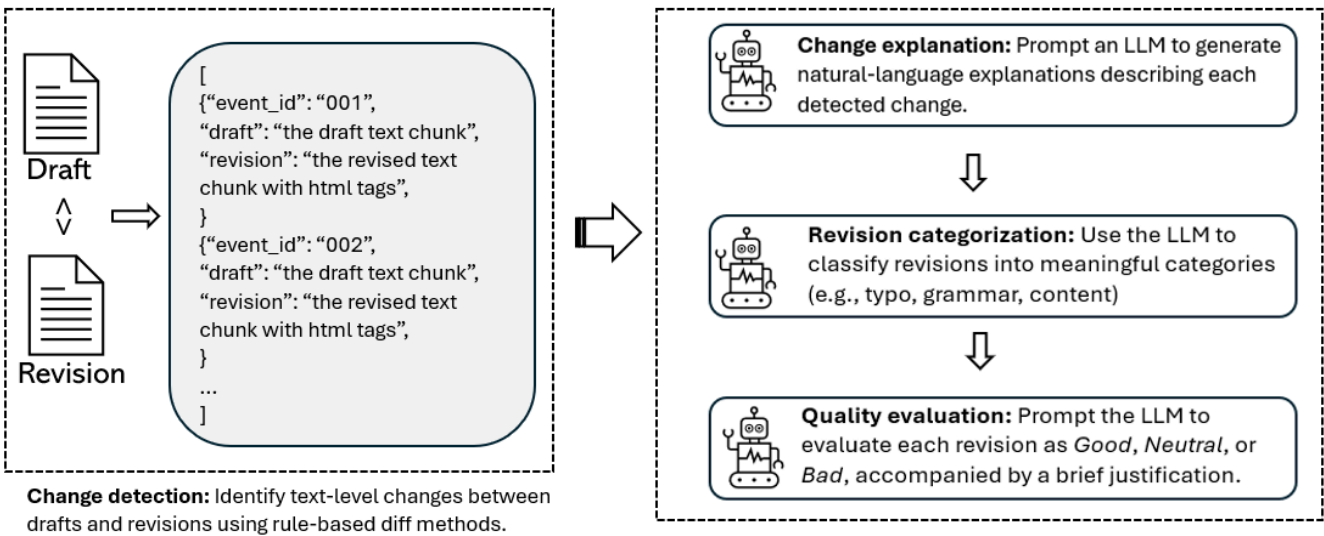
Appendix B

Table B1.

Rubric for Revision Identification and Evaluation

Category	Definition	Decision Rule	Example
Typographic Revision (TYPO)	Changes that correct visual or mechanical surface errors not affecting meaning (e.g., spelling typos, missing/extra characters, accidental punctuation).	If the only difference is a spelling correction, punctuation fix, capitalization, or spacing and the semantic/pragmatic content is unchanged → TYPO.	"teh" → "the"; "its", → "it's"; extra double space removed
Grammatical & Conventions Correction (GRAMMAR)	Changes that fix grammar, morphology, tense/aspect, subject-verb agreement, word order that affect grammaticality but not content meaning substantially.	If the edit improves grammaticality or register (e.g., corrects run-on, reorders words to follow grammar) but does not add new content or change argumentative stance → GRAMMAR.	"He go to school" → "He goes to school"; "in the 1990s" → "during the 1990s"
Word Change (WORD)	Changes in word choice that adjust word-level meaning, specificity, or register without major structural reorganization. Includes synonym replacements, word-level clarifications, or idiom fixes.	If single-word or short phrase swaps change lexical precision, tone, or style, but not intended to change content or add new information → WORD	"increase" → "rise".
Sentence Restructuring (SENTENCE)	Revisions that reorganize clause/sentence structure without substantial addition/removal of content — e.g., splitting/combining sentences, passive→active, reorder of clauses for flow.	If the rewrite changes sentence boundaries or clause order to improve clarity/coherence but does not substantially change content facts → SENTENCE.	"Although it rained he went." → "He went, although it rained."
Content Revision (CONTENT)	Revisions that add, remove, or substantially change content. This is substantive: new ideas, elaboration, addition/removal of examples, thesis change.	If the edit alters meaning, adds new details, removes or changes the information → CONTENT.	"The study was inconclusive." → "The study failed to reach significance due to small sample size, suggesting further research." (addition/clarification).