Checking It Twice: Does Adding Spelling and Grammar Checkers Improve Essay Quality in an Automated Writing Tutor?

Kathryn S. McCarthy[1], Rod D. Roscoe[2], Aaron D. Likens[3], and Danielle S. McNamara[2]

[1]Georgia State University

[2]Arizona State University

[3]University of Nebraska at Omaha

Abstract

This study investigated the effect of incorporating spelling and grammar checking tools within an automated writing tutoring system, Writing Pal. High school students (n = 119) wrote and revised six persuasive essays. After initial drafts, all students received formative feedback about writing strategies. Half of the participants were also given access to spelling and grammar checking tools during the writing and revision periods. Linear mixed effects models revealed that essay quality for students in both conditions improved from initial draft to revision in terms of all aspects except essay unity. The availability of spelling and grammar checking yielded added improvements from initial draft to revision for several aspects of essay quality (i.e., conclusion, organization, voice, grammar/mechanics, and word choice), but other aspects were unaffected (i.e., introduction, body, unity, and sentence structure). The availability of spelling and grammar checking tools had no effect on holistic essay scores. These results indicate that automated spelling and grammar feedback contribute to modest, incremental improvements in writing quality that may complement automated strategy feedback.

Keywords: Automated writing evaluation, Feedback, Natural language processing, Spelling and grammar checking, Writing strategies

Checking It Twice: Does Adding Spelling and Grammar Checkers Improve Essay Quality in an Automated Writing Tutor?

Kathryn S. McCarthy[1(✉)], Rod D. Roscoe[2], Aaron D. Likens[3], and Danielle S. McNamara[2]

[1] Georgia State University, Atlanta, GA, USA
kmccarthyl2@gsu.edu
[2] Arizona State University, Tempe, AZ, USA
{rod.roscoe,dsmcnamal}@asu.edu
[3] University of Nebraska at Omaha, Omaha, NE, USA
alikens@unomaha.edu

Abstract. This study investigated the effect of incorporating spelling and grammar checking tools within an automated writing tutoring system, Writing Pal. High school students (n = 119) wrote and revised six persuasive essays. After initial drafts, all students received formative feedback about writing strategies. Half of the participants were also given access to spelling and grammar checking tools during the writing and revision periods. Linear mixed effects models revealed that essay quality for students in both conditions improved from initial draft to revision in terms of all aspects except essay unity. The availability of spelling and grammar checking yielded added improvements from initial draft to revision for several aspects of essay quality (i.e., conclusion, organization, voice, grammar/mechanics, and word choice), but other aspects were unaffected (i.e., introduction, body, unity, and sentence structure). The availability of spelling and grammar checking tools had no effect on holistic essay scores. These results indicate that automated spelling and grammar feedback contribute to modest, incremental improvements in writing quality that may complement automated strategy feedback.

Keywords: Automated writing evaluation · Feedback · Natural language processing · Spelling and grammar checking · Writing strategies

1 Introduction

National standards report that a majority of students are not proficient writers [1]. One challenge is that the development of writing skills requires repeated, deliberate practice with formative feedback [2], yet providing opportunities for students to compose essays and receive extensive feedback is time-consuming for instructors. Although the U.S. Department of Education recommends at least one hour of writing each day, writing instruction is often overlooked compared to other skills [3, 4]. However, sophisticated natural language processing (NLP) tools have made it possible for automated writing

S. Isotani et al. (Eds.): AIED 2019, LNAI 11625, pp. 270–282, 2019.
https://doi.org/10.1007/978-3-030-23204-7_23

evaluation systems (AWEs) to partially address these challenges. AWEs can provide (a) rapid and accurate evaluations and (b) formative feedback that are otherwise unfeasible in classroom settings constrained by time and resources. Students who receive AWE support have also demonstrated improved writing skills [5–9].

As one example, the Writing Pal (W-Pal [10–13]) is an AWE and intelligent tutoring system (ITS) that targets persuasive writing. Learners are typically given 25 min to write an essay in response to an SAT-like persuasive writing prompt. W-Pal employs NLP-driven algorithms to evaluate the essays [14]. W-Pal algorithms deliver both summative feedback (i.e., a holistic score on a 6-point scale) and formative feedback about writing strategies and improving essay quality. This strategy feedback is related to eight different aspects of writing: freewriting, planning, introduction, body, conclusion, unity, paraphrasing, and revision (see [13]). W-Pal lessons and practice improve students' writing quality and strategy knowledge, and both students and teachers find W-Pal instruction to be valuable and informative [13, 15, 16].

1.1 What About Spelling and Grammar?

The development of technologies that deliver valid automated formative feedback has required innovations in NLP, modeling, and artificial intelligence. Nonetheless, a common question from educators is "Can it check my students' spelling and grammar?" Research suggests that classroom teachers devote extensive time and attention to grammar and mechanics, often at the expense of other writing instruction [17, 18]. Although spelling and grammar checkers have proven useful in ITSs for second language learning (e.g., [19]), the benefits of these tools for writing composition may be questionable. Research suggests that spelling and grammar instruction has little effect on writing quality after early elementary school [20], and spelling and grammar errors are only weakly correlated to expert judgments of essay quality [21]. More effective instruction focuses on writing strategies, such as drafting, organizing, and revising in combination with individualized feedback to improve these strategies (see [22]). Indeed, strategy instruction interventions have yielded large effect sizes [23].

Although spelling and grammar instruction alone may be insufficient for improving students' overall writing skill, this type of feedback may still be valuable. Struggling with spelling can exhaust cognitive resources that would be better put toward other aspects of the writing process [24, 25]. Moreover, in contrast to writing experts, non-expert readers may base their evaluations of quality on spelling and grammar [26–28]. For instance, non-experts use spelling and grammar mistakes to make judgments about a writers' intelligence and other personality traits [29, 30]. Johnson and colleagues [30] asked college students to read writing samples that contained no errors, low-level errors (spelling and grammar mistakes), high-level errors (structural and conceptual errors), or both types of errors. Participants rated the quality of the essays as well as their perceptions of the authors. Low-level errors resulted in more harsh judgments, both in terms of the essay quality and negative personality traits to the writers. In other words, spelling and grammar are still important considerations for writing quality, assessment, and audience perceptions.

1.2 The Current Study

The current study examines the effect of incorporating spelling and grammar checking (SGC) tools into an automated writing evaluation and tutoring system for writing (i.e., W-Pal). W-Pal already provides strategy training and formative feedback for more complex aspects of the writing process (e.g., planning, elaboration, and unity). Thus, this study poses the research question: *In an AWE system, what is the added value of spelling and grammar feedback when combined with higher-level strategy feedback?* In this experiment, all students write essays, receive strategy feedback, and then revise their essays. However, half of the students are also given access to (and reminders to use) spelling and grammar checking tools while writing and revising.

Spelling and grammar feedback could *benefit* students' overall essay quality by helping them focus on content and structure-based writing and revision [24, 25]. Alternatively, spelling and grammar information may have *no effect* on performance. This prediction is driven by the research that indicates essay quality is a function of deeper aspects of the content and style (e.g., [21]). Finally, spelling and grammar feedback could *reduce* essay quality. This outcome might arise because too much prompting and feedback can be distracting rather than helpful (e.g., [31]). Students also often default to "unproductive" superficial revisions [32, 33], and the inclusion of spelling and grammar feedback may misdirect students away from substantive revisions.

2 Method

2.1 Participants

High school students (n = 143) were recruited from a large metropolitan area in the southwestern United States and received financial compensation for their participation. Twenty-four participants were omitted due to incomplete data as a result of either technical or experimenter errors, resulting in a total of 119 participants.

One student completed the study but did not provide demographic information. Thus, the sample (M_{age} = 17.19, SD = 1.28, Range: 13–19) was 61.3% female and 37.5% male; 53.8% Caucasian, 21% Hispanic (Latin American), 10.1% Asian, 7.6% African American, and 6.7% reported as other. Finally, 85.7% of participants were native English speakers.

2.2 Design and Procedure

Students were randomly assigned one of two conditions in which they received either writing strategy feedback (Strategy Condition, n = 60) or writing strategy feedback along with spelling and grammar checking tools (Strategy + SGC Condition, n = 59).

The study included four sessions. In the first, session, students provided demographic information and then completed the Gates-MacGinitie Reading Test (GRMT; [34]) as a measure of reading skill. In each of the subsequent three sessions, the students wrote and revised two essays on different persuasive writing prompts

(Table 1). Each of the prompts had a brief introduction to frame the issue and ended in a question. For example, the prompt about images and impressions read:

> All around us appearances are mistaken for reality. Clever advertisements create favorable impressions but say little or nothing about the products they promote. In stores, colorful packages are often better than their contents. In the media, how certain entertainers, politicians, and other public figures appear is sometimes considered more important than their abilities. All too often, what we think we see becomes far more important than what really is. Do images and impressions have a positive or negative effect on people?

Table 1. Essay prompt questions in order of presentation

Prompt question
1. Do images and impressions have a positive or negative effect on people?
2. Do people achieve more success by cooperation or by competition?
3. Do people place too much emphasis on winning?
4. Should people always maintain their loyalties, or is it sometimes necessary to switch sides?
5. Is it better for people to act quickly and expect quick responses from others rather than to wait patiently for what they want?
6. Do personal memories hinder or help people in their effort to learn from their past and succeed in the present?

Participants were allotted 25 min to write an initial draft. After this time elapsed, W-Pal assigned a holistic score from "Poor" (1) to "Great" (6) along with individual strategy feedback aligned with W-Pal strategy lessons (e.g., planning, paragraph quality, unity, and paraphrasing). After viewing this feedback, participants were given 10 min to revise. Feedback messages provide actionable steps to help the participant improve their essay. For example, if the algorithm determined that the essay was too short, the participant might receive the following strategy feedback message:

> This essay may not have enough paragraphs to fully support the main argument. If you need help developing support for future drafts or essays, it may be helpful to freewrite.
>
> - Write down possible arguments that may relate to your thesis
> - Brainstorm as many relevant facts and examples as you can
> - Try to think of details from school classes, news stories, and your own life that may relate to the arguments!

For participants in the Strategy + SGC Condition, "Check Spelling" and "Check Grammar" buttons appeared at the bottom of the interface during the writing and revision periods (Fig. 1). Participants could access either function at any time while writing or revising. Errors were detected using the open source API LanguageTool [35, 36]. When these tools were selected, errors were underlined similar to common word processors. Clicking on the error opened a small pop-up window with potential corrections. A reminder about the SGC tools appeared when there were 5 min remaining in the writing session, but students were not forced to use the tools.

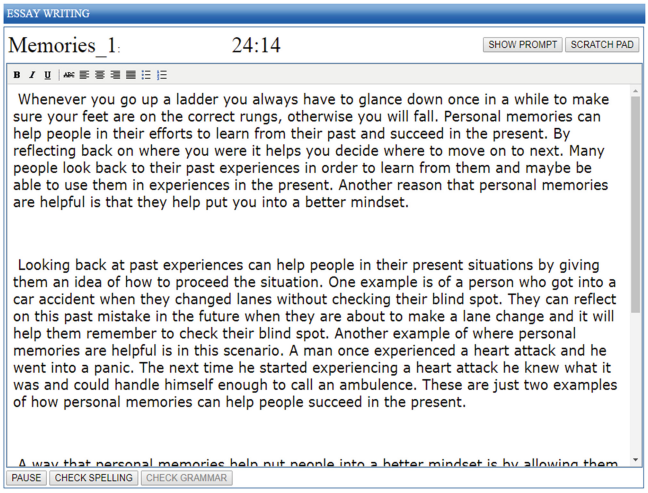

Fig. 1. W-Pal writing window with spelling and grammar checkers

2.3 Writing Assessment and Scoring

In the study, W-Pal algorithms determined immediate scores and feedback. For later analyses, human raters (following the same criteria underlying the algorithm) assigned holistic scores and nine subscores (i.e., specific writing traits) on both drafts. Raters were four graduate students of English. Rater pairs were trained to a high level of reliability (all kappas > .80) on all metrics. Holistic scores (1–6 scale) were based on the standard SAT rubric and subscores (1–6 scale) aligned with W-Pal lessons:

- **Grammar, Style, & Mechanics:** essay conveys strong control of the standard conventions of writing; avoiding errors in grammar, syntax, and mechanics
- **Introduction:** writer demonstrates mastery in meeting the goals of an introduction (e.g., presenting a topic, providing a purpose, clearly stating a thesis, and previewing arguments)
- **Body:** writer demonstrates mastery in meeting the goals of body arguments (e.g., transition between arguments, using topic sentences, supporting arguments with evidence, and maintaining a flow throughout the arguments)
- **Conclusion:** writer demonstrates mastery in meeting the goals of a conclusion (e.g., summarizing the essay, re-establishing the significance of discussion, capturing the reader's attention, and effectively closing the essay)
- **Organization:** essay follows a logical structure (e.g., introduction, body arguments and evidence, conclusion)
- **Unity:** details support the thesis and do not stray from the prompt and the main ideas and organizing principles presented in the introduction
- **Voice:** writer is expressive, engaging, and sincere; a strong sense of audience
- **Word Choice:** writer is precise and effective in word choice
- **Sentence Structure:** sentence patterns are varied effectively, enhancing the quality of the essay

Note that holistic scores was not calculated from the subscales (e.g., an average)—holistic scores and subscores were distinct judgments although likely to be correlated.

Each essay (N = 1435) was scored by two raters. Across raters, the reliability of ratings for scores ranged from ICC = .79 to .90. The final scores for each essay reflect the average score of the two raters.

3 Results

Table 2 presents average scores as a function of draft and feedback condition. Overall, participants exhibited modest yet significant increases in average essay scores from initial draft (M = 3.63, SD = .55) to revision (M = 3.76, SD = .53), $F(1, 117) = 25.44$, $p < .000$, $\eta^2 = .18$. Holistic writing scores were strongly correlated with all writing subscores (r = .73 to .93), supporting the concurrent validity of the assessment. Consistent with existing research [37], reading skill (i.e., GMRT scores) was also positively correlated with holistic score and subscores (r = .47 to .68).

Table 2. Means and standard deviations of essay scores as a function of experimental condition and essay draft

Score (1–6)	Strategy condition		Strategy + SGC condition	
	Initial	Revised	Initial	Revised
Holistic	3.49 (0.66)	3.64 (0.64)	3.38 (0.69)	3.52 (0.70)
Grammar, Mechanics*	3.55 (0.61)	3.57 (0.56)	3.51 (0.60)	3.69 (0.60)
Introduction	3.83 (0.69)	3.93 (0.62)	3.66 (0.65)	3.76 (0.55)
Body	3.77 (0.60)	3.82 (0.55)	3.64 (0.65)	3.75 (0.62)
Conclusion*	3.43 (0.74)	3.52 (0.72)	2.98 (0.83)	3.35 (0.77)
Organization*	3.83 (0.57)	3.93 (0.55)	3.53 (0.60)	3.77 (0.66)
Unity	4.13 (0.52)	4.10 (0.49)	3.94 (0.49)	4.01 (0.50)
Voice*	3.77 (0.54)	3.85 (0.54)	3.61 (0.58)	3.90 (0.53)
Word choice*	3.82 (0.56)	3.89 (0.47)	3.67 (0.54)	3.88 (0.57)
Sentence structure	3.59 (0.54)	3.73 (0.55)	3.53 (0.58)	3.59 (0.60)

Note. An asterisk (*) indicates that the Strategy + SGC condition demonstrated statistically significant gains from initial to revised whereas those in the Strategy Condition did not.

3.1 Analysis of Essay Improvement and Condition

Linear mixed effects (LME) models were conducted to detect the effects of the spelling and grammar checker tools on both initial draft and revision while also accounting for potential influences of the essay prompt or reading skill.

Holistic writing score and all nine writing subscores were fit with the same series of LME models using the lme4 package [38] in R [39]. Simple slopes were estimated using the *reghelper* package [40]. Table 3 shows the variables entered into each model.

For the *Baseline* model (M0), only GMRT scores and a prompt factor were included. Second, a *Draft* model (M1) included a draft factor (i.e., initial vs. revised draft) to investigate how scores changes as a function of revising. Finally, a *Condition* model (M2) added a condition factor (i.e., Strategy vs. Strategy + SG) along with two interaction terms to assess the effect of condition. The interaction terms included "Draft × Condition" and "GMRT × Condition". Likelihood ratio tests were used to compare model fit. Significant chi-square (χ^2) tests indicate that adding the additional variable(s) improved fit as compared to the previous model (Table 4). These analyses revealed mixed results for the added value of SGC tools in an AWE. Benefits were observed for several subscores but not holistic quality.

Table 3. Linear mixed effects model design

Model	Variables included
Baseline (M0)	GMRT + Prompt
Draft (M1)	GMRT + Prompt + Draft
Condition (M2)	GMRT + Prompt + Draft + Condition + (Draft × Condition) + (GMRT × Condition)

Table 4. Likelihood ratio tests (χ^2) comparing fit with additional variables

	Draft (M1) $\chi^2(1)$	Condition (M2) $\chi^2(3)$
Holistic	24.15***	*ns*
Grammar, Mechanics	13.87***	15.06***
Introduction	24.15***	*ns*
Body	6.56*	*ns*
Conclusion	26.68***	14.01***
Organization	20.34***	7.42+
Unity	*ns*	*ns*
Voice	27.02***	11.56**
Word choice	26.04***	9.08*
Sentence structure	9.59**	*ns*

Note. *ns* = not significant; ***$p < .001$; **$p < .01$; *$p < .05$; $^{+}p < .06$.

3.2 Effects on Holistic Writing Quality and Grammar, Style, and Mechanics

One overarching question was whether the accessibility of SGC tools influenced overall writing quality. The LME models suggest that this was not the case. With regards to *Holistic Score*, the Draft model demonstrated improved model fit compared to the Baseline model. However, the Condition model did not further improve model fit. Although students improved their essays through revising, the availability of spelling and grammar checking tools did not appear to significantly contribute to holistic gains.

If students receive valid and useful feedback and writing mechanics, these benefits should influence *Grammar, Style, and Mechanics* subscores, even if holistic quality was not affected. Indeed, this was the observed pattern. The Draft model exhibited improved model fit compared to Baseline. Further, the Condition model demonstrated further improved model fit. Simple slopes revealed that Strategy + SGC Condition participants improved from initial draft to revision, Estimate = 0.26, SE = 0.04, $t(1283) = 4.86$, $p < 0.001$, whereas participants in the Strategy Condition did not, Estimate = 0.02, SE = 0.04, $t(1283) = 0.50$, $p = 0.62$. The availability of spelling and grammar feedback had little effect on initial drafts, but facilitated revising with respect to grammar, style, and mechanics, as should be expected. In broader terms, spelling and grammar tools contributed to incremental improvements in writing quality that were not necessarily reflected in holistic essay quality.

3.3 Additional Benefits of Spelling and Grammar Feedback

Several other subscores provided evidence that the availability of spelling and grammar tools facilitated incremental gains in specific aspects of writing. The quality of *Conclusion*, *Organization, Voice*, and *Word Choice* significantly improved from the initial draft to the revision for students in the Strategy + SGC Condition, whereas there was no improvement for students in the Strategy Condition. These findings are reflected within the LME by improved model fit compared to baseline in the Draft model, and improved fit of the Condition model, specifically driven by the Draft × Condition interaction term. Simple slopes—with draft as the focal predictor and condition as the moderator—revealed that participants who had access to spelling and grammar tools tended to increase in subscores from draft to revision (*Conclusion* Estimate = 0.38 SE = 0.06, $t(1283) = 6.07$, $p < 0.001$; *Organization* Estimate = 0.23, SE = 0.05, $t(1283) = 4.86$, $p < 0.001$; *Voice* Estimate = 0.29, SE = 0.05, $t(1283) = 5.97$, $p < 0.001$; *Word Choice* Estimate = 0.22, SE = 0.04, $t(1283) = 5.54$, $p < 0.001$), whereas there was no observed improvements for those in the Strategy Condition (all Estimates < .09; $t < 2$).

3.4 Writing Subscores Unaffected by Spelling and Grammar Feedback

As hinted by the lack of effects on holistic writing quality, the benefits of SGC tools were not universal—several writing subscores showed no effect of feedback condition. These included *Introduction*, *Body*, and *Sentence Structure*. Participants improved on these traits from initial to revised draft (i.e., revising improved the essay) but condition (i.e., adding spelling and grammar feedback) had no influence (see Table 4).

Finally, contrary to other subscores, participants did not show any change in *Essay Unity* scores across drafts; the Draft model did not improve model fit compared to Baseline. The Condition model also did not improve model fit.

4 Discussion and Future Work

Computer-based educational tools can provide a wealth of feedback to student writers, including summative scores on writing quality and individualized, formative feedback on strategies and ways to improve [41]. Similar to word processing programs, these tools might also incorporate spelling and grammar checking. However, one question is whether having access to SGC tools benefits students above and beyond strategy feedback. Prior research suggests that an overemphasis on writing mechanics can be useless or detrimental [21], whereas strategy instruction and feedback are beneficial [22, 23].

To explore these questions, the current study examined how the inclusion of spelling and grammar feedback in an AWE system affected the quality of essays and revisions. This study focused on high school students, a critical target population for writing instruction and intervention. Analyses examined gains in holistic essay scores and nine subscores (rated by writing experts) as a function of feedback condition.

4.1 Incremental Improvements in Essay Quality

Linear mixed effect models indicated that essay quality improved holistically and along all subscores (except unity) as a function of revising with feedback. When students wrote essays in W-Pal, received feedback, and revised, their essays improved. This finding replicates previous results showing that strategy feedback results in improvements from initial draft to the revision [42–45].

As one might expect, the availability of checking tools improved the grammar and mechanics in the essays. However, spelling and grammar feedback, in conjunction with strategy feedback, *also* improved essays from initial draft to revision on the dimensions of conclusion paragraphs, organization, voice, and word choice. This finding is important because it suggests that providing students with tools to check their spelling and grammar might (a) free up resources to consider other aspects of writing when writing and/or (b) inspire a greater willingness to revise. Grammar checkers have been shown to increase students' motivation and confidence in their writing [46, 47]. Thus, spelling and grammar checkers may have direct benefits on mechanics that then afford indirect benefits elsewhere. Future research should consider students' subjective reactions to these tools, such as whether they indeed perceive writing to be less burdensome or more engaging when they have a "safety net" of checking tools in an AWE system.

Importantly, these tools did *not* appear to benefit holistic essay scores. This is consistent with prior work showing that expert evaluations of essay quality rely on deeper features of text [21]. Likewise, there were no apparent benefits of spelling and grammar feedback on introduction quality, body quality, unity, or sentence structure. Thus, although the tools were moderately useful, they were not universally beneficial.

Finally, it is worth noting that this feedback did not *reduce* performance on any subscores. Overall, these findings indicate that adding spelling and grammar checkers *in conjunction with strategy feedback* is moderately beneficial for AWEs.

4.2 Directions for Future Research

In the current study, participants did not improve the unity of their essays, regardless of feedback or checking tools. One explanation may be that unity requires the writer to step back and evaluate the connectivity, or cohesion, of the essay as a whole. A 10-minute window for revision may not provide sufficient time to conceptualize and implement such revisions. An alternative explanation could be that participants received fewer messages about essay unity than other feedback topics. Further analyses of the types of feedback that participants received, and their viewing of that feedback, may shed further light on how we might help students to improve the unity of their essays.

In addition to assessing the specific feedback messages received and viewed, future work might also analyze how students approached the revising task (i.e., the frequency and type of edits made). One possibility is that students in the strategy feedback condition used their 10 min of revision time to address mechanical errors—a typical bias toward proofreading over substantive revising. In contrast, because students in the spelling and grammar condition had already completed this task (at least partially), they spent more time during revising on substantive edits. Given that spelling and grammar feedback did not impact the subscores equally, these findings may shed light on which aspects of writing, or which writing strategies, are prioritized by students. Eliciting students' self-reported rationale for revising or using AWE tools could help to understand how they navigate AWE functions and uptake feedback. For instance, students might focus on the "easiest" or "fastest" edits, or they may prioritize the most "critical" flaws.

Finally, having these tools available might change the dynamics of the writing process [48, 49]—students may write *more* or *faster* when they feel they can rely on SGC tools to make the task easier. To explore these plausible changes in writing production, the use of log-files, key strokes, and similar data may help to elucidate how students use the spelling and grammar functions, and the extent that tool use—as opposed to mere availability or accessibility—influences drafting and revising activities. The benefits of using spelling and grammar tools are likely to be more nuanced than a simple "more is better" assumption. Students might rely on the tools to mechanically and mindlessly fix typos instead of reflecting on the meaning of their writing or the possible reactions of their audience.

4.3 Conclusion

Questions about optimal feedback—including feedback content, timing, methods, and effects on performance—are among the most critical challenges facing educators, researchers, or others who develop and implement adaptive educational technologies. A number of nuanced factors influence feedback quality and students' feedback uptake.

This study explored feedback in the context of essay writing with AWE support. Research has previously established that formative feedback on writing strategies is effective whereas feedback that exclusively targets grammar and mechanics is not effective. Nonetheless, teachers, students, and writers intuitively crave feedback and automated corrections on these writing features. This study provides compelling evidence that students benefit from both types of feedback, and that guidance on writing mechanics does not inhibit writing quality or deter from revising.

Several questions remain concerning the source, loci, and dynamics of these effects. Do benefits stem from helping students manage their resources and focus on rhetorical aspects of writing? Do benefits stem from a sense of writer empowerment? How do varying feedback tools influence real-time writing and revising behaviors? Ultimately, the objective is to enhance students' ability to improve their writing, and automated feedback affords multiple resources for accomplishing that goal.

Acknowledgements. This research was supported by grants from the U.S. DoEd Institute of Education Sciences (R305A120707 and R305A180261) and the U.S. DoD Office of Naval Research (N000141712300). Opinions, findings, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding sources.

References

1. National Center for Education Statistics: The Nation's report Card: Washington, D.C.: Institute of Education Sciences, U.S. Department of Education (2012)
2. Kellogg, R.T., Raulerson III, B.A.: Improving the writing skills of college students. Psychon. Bull. Rev. **14**(2), 237–242 (2007)
3. Graham, S., Capizzi, A., Harris, K.R., Hebert, M., Morphy, P.: Teaching writing to middle school students: a national survey. Read. Writ. **27**(6), 1015–1042 (2014)
4. Kiuhara, S.A., Graham, S., Hawken, L.S.: Teaching writing to high school students: a national survey. J. Educ. Psychol. **101**(1), 136–160 (2009)
5. Allen, L.K., Jacovina, M.E., McNamara, D.S.: Computer-based writing instruction. In: MacArthur, C.A., Graham, S., Fitzgerald, J. (eds.) Handbook of Writing Research, 2nd edn, pp. 316–329. The Guilford Press, New York (2016)
6. Palermo, C., Thomson, M.M.: Teacher implementation of self-regulated strategy development with an automated writing evaluation system: effects on the argumentative writing performance of middle school students. Contemp. Educ. Psychol. **54**, 255–270 (2018)
7. Shermis, M.D., Burstein, J. (eds.): Handbook of Automated Essay Evaluation: Current Applications and New Directions. Routledge, New York (2013)
8. Stevenson, M., Phakiti, A.: The effects of computer-generated feedback on the quality of writing. Assess. Writ. **19**, 51–56 (2014)
9. Wilson, J., Czik, A.: Automated essay evaluation software in English language arts classrooms: effects on teacher feedback, student motivation, and writing quality. Comput. Educ. **100**, 94–109 (2016)
10. Crossley, S.A., Allen, L.K., McNamara, D.S.: The Writing Pal: a writing strategy tutor. In: Crossley, S.A., McNamara, D.S. (eds.) Adaptive educational technologies for literacy instruction, pp. 204–224. Routledge, New York (2016)
11. Dai, J., Raine, R.B., Roscoe, R., Cai, Z., McNamara, D.S.: The writing-pal tutoring system: development and design. J. Eng. Comput. Innov. **2**, 1–11 (2011)
12. McNamara, D.S., et al.: The writing-pal: natural language algorithms to support intelligent tutoring on writing strategies. In: McCarthy, P.M., Boonthum-Denecke, C. (eds.) Applied Natural Language Processing and Content Analysis: Identification, Investigation, and Resolution, pp. 298–311. IGI Global, Hershey (2012)
13. Roscoe, R.D., McNamara, D.S.: Writing Pal: feasibility of an intelligent writing strategy tutor in the high school classroom. J. Educ. Psychol. **105**, 1010–1025 (2013)
14. McNamara, D.S., Crossley, S.A., Roscoe, R.D., Allen, L.K., Dai, J.: A hierarchical classification approach to automated essay scoring. Assess. Writ. **23**, 35–59 (2015)

15. Allen, L.K., Crossley, S.A., Snow, E.L., McNamara, D.S.: Game-based writing strategy tutoring for second language learners: game enjoyment as a key to engagement. Lang. Learn. Technol. **18**, 124–150 (2014)
16. Roscoe, R.D., Allen, L.K., Weston, J.L., Crossley, S.A., McNamara, D.S.: The Writing Pal intelligent tutoring system: usability testing and development. Comput. Compos. **34**, 39–59 (2014)
17. Cutler, L., Graham, S.: Primary grade writing instruction: a national survey. J. Educ. Psychol. **100**(4), 907–919 (2008)
18. Morris, D., Blanton, L., Blanton, W.E., Perney, J.: Spelling instruction and achievement in six classrooms. Elem. Sch. J. **96**(2), 145–162 (1995)
19. Heift, T., Rimrott, A.: Learner responses to corrective feedback for spelling errors in CALL. System **36**(2), 196–213 (2008)
20. Graham, S., Perin, D.: A meta-analysis of writing instruction for adolescent students. J. Educ. Psychol. **99**(3), 445–476 (2007)
21. Crossley, S.A., Kyle, K., Allen, L.K., McNamara, D.S.: The importance of grammar and mechanics in writing assessment and instruction: evidence from data mining. In: Stamper, J., Pardos, Z., Mavrikis, M., McLaren,B.M. (eds.) Proceedings of the 7th International Conference on Educational Data Mining, London, UK, pp. 300–303 (2014)
22. Graham, S., Harris, K.R., Chambers, A.B.: Evidence-based practice and writing instruction. Handb. Writ. Res. **2**, 211–226 (2016)
23. Graham, S., Harris, K.R.: Strategy instruction and the teaching of writing. Handb. Writ. Res. **5**, 187–207 (2006)
24. Graham, S., Santangelo, T.: Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. Read. Writ. **27**(9), 1703–1743 (2014)
25. Morphy, P., Graham, S.: Word processing programs and weaker writers/readers: a meta-analysis of research findings. Read. Writ. **25**(3), 641–678 (2012)
26. Marshall, J.C.: Composition errors and essay examination grades re-examined. Am. Educ. Res. Assoc. **4**, 375–385 (1967)
27. Marshall, J.C., Powers, J.M.: Writing neatness, composition errors, and essay grades. J. Educ. Meas. **6**, 97–101 (1969)
28. Figuredo, L., Varnhagen, C.K.: Didn't you run the spell checker? Effects of type of spelling error and use of a spell checker on perceptions of the authors. Read. Psychol. **26**, 441–458 (2005)
29. Boland, J.E., Queen, R.: If you're house is still available, send me an email: personality influences reactions to written errors in email messages. PLoS ONE **11**, e0149885 (2016)
30. Johnson, A.C., Wilson, J., Roscoe, R.D.: College student perceptions of writing errors, text quality, and author characteristics. Assess. Writ. **34**, 72–87 (2017)
31. McCarthy, K.S., Likens, A.D., Johnson, A.M., Guerrero, T.A., McNamara, D.S.: Metacognitive overload!: Positive and negative effects of metacognitive prompts in an intelligent tutoring system. Int. J. Artif. Intell. Educ. **28**(3), 1–19 (2018)
32. Crawford, L., Lloyd, S., Knoth, K.: Analysis of student revisions on a state writing test. Assess. Eff. Interv. **33**, 108–119 (2008)
33. Fitzgerald, J.: Research on revision in writing. Rev. Educ. Res. **57**, 481–506 (1987)
34. MacGinitie, W.H., MacGinitie, R.K., Cooter, R.B., Curry, S.: Assessment: Gates-MacGinitie reading tests. Read. Teach. **43**(3), 256–258 (1989)
35. Miłkowski, M.: Developing an open-source, rule-based proofreading tool. Softw. Pract. Exp. **40**(7), 543–566 (2010)
36. Naber, D.: A rule-based style and grammar checker. Master's thesis, Universität Bielefeld (2003). http://www.danielnaber.de/publications

37. Allen, L.K., Snow, E.L., Crossley, S.A., Jackson, G.T., McNamara, D.S.: Reading comprehension components and their relation to the writing process. L'année psychologique/Top. Cogn. Psychol. **114**, 663–691 (2014)
38. Bates, D., Maechler, M., Bolker, B., Walker, S.: lme4: linear mixed-effects models using Eigen and S4. R package version 1.1–7 (2015)
39. R Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2017)
40. Hughes, J., Team, R.C. Reghelper: helper functions for regression analysis. R package, version 0.3.3 (2017)
41. Strobl, C., et al.: Digital support for academic writing: a review of technologies and pedagogies. Comput. Educ. **131**, 33–48 (2018)
42. Attali, Y., Burstein, J.: Automated essay scoring with e-rater® v. 2.0. ETS Research Report Series 2004. 2, i-21(2004)
43. Crossley, S.A., Roscoe, R.D., McNamara, D.S.: Using automatic scoring models to detect changes in student writing in an intelligent tutoring system. In: Proceedings of the 26th International Florida Artificial Intelligence Research Society (FLAIRS) Conference, pp. 208–213. The AAAI Press, Menlo Park (2013)
44. Roscoe, R.D., Snow, E.L., McNamara, D.S.: Feedback and revising in an intelligent tutoring system for writing strategies. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 259–268. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_27
45. Roscoe, R.D., Snow, E.L., Allen, L.K., McNamara, D.S.: Automated detection of essay revising patterns: application for intelligent feedback in a writing tutor. Technol. Instr. Cogn. Learn. **10**(1), 59–79 (2015)
46. Cavaleri, M.R., Dianati, S.: You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. Journal of Academic Language and Learning **10**(1), A223–A236 (2016)
47. Potter, R., Fuller, D.: My new teaching partner? Using the grammar checker in writing instruction. Engl. J. **98**(1), 36–41 (2008)
48. Allen, L.K., et al.: {ENTER}ing the time series {SPACE}: Uncovering the writing process through keystroke analysis. In: Proceedings of the 9th International Conference on Educational Data Mining (EDM 2016), pp. 22–29. International Educational Data Mining Society, Raleigh, NC (2016)
49. Likens, A.D., Allen, L.K., McNamara, D. S.: Keystroke dynamics predict essay quality. In: Gunzelmann, G., Howes, A., Tenbrink, T., Davelaar, E. (eds.) Proceedings of the 39th Annual Meeting of the Cognitive Science Society (CogSci 2017), pp. 2573–2578. Cognitive Science Society, London, UK (2017)