Am I Wrong or Am I Right? Gains in Monitoring Accuracy in an Intelligent Tutoring System for Writing

Laura K. Allen , Scott A. Crossley, Erica L. Snow, Matthew E. Jacovina, Cecile Perret, and Danielle S. McNamara

Am I Wrong or Am I Right? Gains in Monitoring Accuracy in an Intelligent Tutoring System for Writing

Laura K. Allen[1], Scott A. Crossley[2], Erica L. Snow[1], Matthew E. Jacovina[1], Cecile Perret[1], and Danielle S. McNamara[1]

[1]Learning Sciences Institute, Arizona State University, Tempe, AZ 85287
{LauraKAllen, Erica.L.Snow, Matthew.Jacovina, CPerret,
DSMcnama}@asu.edu

[2]Department of Applied Linguistics/ESL, Georgia State University, 25 Park Place, Atlanta, GA 30303, USA
scrossley@gsu.edu

Abstract. We investigated whether students increased their self-assessment accuracy and essay scores over the course of an intervention with a writing strategy intelligent tutoring system, W-Pal. Results indicate that students were able to learn from W-Pal, and that the combination of strategy instruction, game-based practice, and holistic essay-based practice led to equivalent gains in self-assessment accuracy compared to heavier doses of deliberate writing practice (offering twice the amount of system feedback).

Keywords: Tutoring, Intelligent Tutoring Systems, Self-Assessment, Metacognition, Writing, Automated Writing Evaluation

1 Introduction

Computer-based writing instruction provides students with feedback on their essays in the absence of a teacher. Research on these instructional systems has largely focused on evaluating the accuracy of the automated scores [1-2], as well as whether students increase the quality of their essays after receiving feedback [3]. Few studies, however, have investigated the impact of these systems on students' ability to monitor their own performance. This is a significant exclusion, because *monitoring accuracy* is important for durable, long-term learning [4]. Unfortunately, students struggle with this skill, indicated by the fact that they are often largely inaccurate in their self-assessments of academic performance [5].

The Writing Pal (W-Pal) is an intelligent tutoring system (ITS) designed to improve the writing proficiency of students through explicit strategy instruction, deliberate practice, and automated feedback [6-7]. Within W-Pal, students are provided with strategy instruction and practice in the context of eight instructional modules, which contain lesson videos and mini-games. Additionally, W-Pal contains an essay-writing component where students can practice holistic essay writing. This feature contains a

word processor where students can generate essays and receive automated summative and formative feedback. Previous studies point to the effectiveness of W-Pal, as training has been linked to gains in essay scores and strategy knowledge over time [7-8].

The purpose of this study is to investigate the efficacy of W-Pal to improve the monitoring accuracy of its student users. Our research questions are outlined below:

1) Prior to writing strategy training, do students provide accurate assessments of their own writing?
2) Does the alignment between the students' self-assessments and the ratings provided by the W-Pal tutor increase over the training sessions?
3) Does the student-system alignment vary according to the type of training that students receive?

2 Method

High school students (n = 87) attended a 10-session study and were randomly assigned to one of two conditions: *W-Pal condition* (n=42) or *Essay condition* (n=45). *Sessions 1 and 2* were devoted to the pretest and posttest, respectively. *Sessions 2-9* were reserved for training. Students in both the W-Pal and Essay conditions began each session by writing and revising one 25-minute essay. Once this draft was complete, they rated their essay, received W-Pal feedback, and were given 10 minutes to revise the essay. Students in the Essay condition then repeated this process (wrote a second essay, self-assessed, received feedback, and revised this essay). Students in the W-Pal condition completed one instructional module (lesson videos and mini-games)

3 Results

W-Pal essay ratings (possible range=1-6) were calculated using the *W-Pal algorithm* (see 2 for details). This score aligns well with expert and teacher ratings of essays [2]. Additionally, students' self-assessments (possible range=1-6) were collected. A *misalignment score* was calculated for each student by taking the absolute value of the difference between the student's self-assessment and the W-Pal essay rating.

3.1 Initial Essay Attempt

On Session 2, all students wrote and self-assessed an essay before receiving feedback. Because students received no training prior to producing this essay, its quality and the self-assessments served as baseline measures of students' abilities. On average, W-Pal assigned these essays a score of 2.35 (*SD*=0.91), whereas students provided an average self-assessment of 3.75 (*SD*=0.89). Thus, in relation to W-Pal, students tended to overestimate their essay ratings; $t(84)=11.36$, $p<.001$. Additionally, the W-Pal and student ratings were not significantly correlated ($r=.20$, $p=.069$). The absence of a significant correlation and the differences in the average ratings are indicative of a weakness in students' monitoring accuracy.

3.2   Alignment during Training

Three repeated-measures ANOVAs were calculated to investigate whether essay scores, self-assessments, and misalignment scores changed across training sessions. Additionally, 8 t-tests were conducted to determine whether misalignment persisted for all sessions. We hypothesized that W-Pal training would lead to an increase in essay scores, but a decrease in self-assessment (to account for overestimation early in training) and misalignment scores

The results support our hypotheses. There was a significant linear effect of essay, self-assessment, and misalignment scores across sessions. Essay scores increased, $F(1,78)=6.31$, $p=.01$, whereas self-assessment [$F(1,81)=28.11$, $p<.001$] and misalignment scores [$F(1,78)=6.49$, $p=.01$] decreased, suggesting that training promoted better alignment between self-assessments and system scores. Results of the t-tests, however, indicated that there were significant differences between scores across all sessions ($p<.001$). Therefore, students' monitoring accuracy still had room for improvement. An important note is that students did not simply perceive their performance to be decreasing across time. A repeated-measures ANOVA on students' responses to a daily survey indicated that students' *perceived writing improvement* increased across the sessions, $F(1,74)=23.57$, $p<.001$.

3.3   Alignment By Training Condition

Our final research question concerned the influence of condition on students' alignment with W-Pal. A mixed-design ANOVA on misalignment scores (session as within-subjects factor; condition as between-subjects factor) indicated that, although there was a significant linear effect of session [$F(1,78)=6.49$, $p=.01$], there was no significant effect of condition ($F<1$), nor an interaction between condition and session ($F<1$).

4   Discussion

Results of this study indicate that students were able to learn from W-Pal, and that the combination of strategy instruction, game-based practice, and essay-writing practice led to equivalent gains in self-assessment accuracy compared to heavier doses of deliberate writing practice (offering twice the amount of feedback). Our interpretation is that students' exposure to writing strategies helped them to increase the accuracy of their performance monitoring. Prior to receiving training, students in this study were largely inaccurate in their self-assessments of essay quality. However, over the course of 8 training sessions, students in both conditions were able to significantly increase the accuracy of these assessments. This interpretation is additionally supported by the similarities found between the training conditions. We suggest that the strategy instruction and game-based practice in the W-Pal condition provided students with a deep understanding of the system feedback, which helped them to understand when they were (and were not) meeting the requirements of the writing task. As a result, these students were able to align their self-assessments with the assessments provided

by the tutor at the same rate as their peers, despite engaging in fewer self-assessments and being exposed to a significantly smaller number of feedback messages.

These results are important because they indicate that computer-based writing instruction can promote better monitoring accuracy amongst students, which is an important element of transfer. In particular, this study suggests that students may not simply be relying on the tutor to provide them with assessments of their own performance. Rather, they seem to be internalizing the information in the feedback and using this to adjust their metacognition over time. Previous research indicates that students' self-assessments are typically inaccurate – therefore, this work has important educational implications, as it suggests that these self-assessments can be enhanced through training with a writing-based tutoring system.

5   Acknowledgments

6   References

1. Attali, Y., Burstein, J.: Automated Essay Scoring with E-rater V.2. Journal of Technology, Learning, and Assessment. 4, 3 (2006)
2. McNamara, D.S., Crossley, S.A., Roscoe, R.D., Allen, L.K., Dai, J. Natural Language Processing in a Writing Strategy Tutoring System: Hierarchical Classification Approach to Automated Essay Scoring. Assessing Writing. 23, 35-59 (2015)
3. 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. Technology, Instruction, Cognition, and Learning (in press)
4. Dunlosky, J., Hertzog, C., Kennedy, M., Thiede, K.: The Self-Monitoring Approach for Effective Learning. Cognitive Technology. 10, 4–11 (2005)
5. Varner L.K., Roscoe, R.D., McNamara, D.S.: Evaluative Misalignment of 10th-Grade Student and Teacher Criteria for Essay Quality: An Automated Textual Analysis. Journal of Writing Research. 5, 35-59 (2013)
6. Roscoe, R., Allen, L., Weston, J., Crossley, S., McNamara, D.: The Writing Pal intelligent tutoring system: Usability testing and development. Computers and Composition. 34, 39-59 (2014)
7. 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. Language Learning and Technology. 18, 124-150 (2014)
8. Crossley, S.A., Varner, L.K., Roscoe, R.D., McNamara, D.S.: Using Automated Cohesion Indices as a Measure of Writing Growth in Intelligent Tutoring Systems and Automated Essay Writing Systems. In K. Yacef et al. (eds.): 16th International Conference on Artificial Intelligence in Education (AIED), pp. 269-278. Heidelberg, Berlin: Springer (2013)