
RUNNING HEAD: Monitoring and Writing

Self-Monitoring Support in Learning to Write

Kwangsu Cho
University of Missouri, Columbia

Moon-Heum Cho
University of Missouri, Columbia

Douglas J. Hacker
University of Utah, Salt Lake City
Abstract

This study examined the role of self-monitoring support for writing skill improvement in a reciprocal peer review of writing system called SWoRD (Cho & Schunn, 2007). Because self-monitoring has been identified as a critical component in self-regulated learning (Butler & Winne, 1995; Cresswell, 2000; Zimmerman & Risemberg, 1997), students were provided with opportunities to self-monitor their writing through self-evaluations and peer evaluations on both their own and peer writing. With 601 graduate and undergraduate students from 16 courses in three U.S. universities, we found that although not all the students developed successful self-monitoring skills, the students who did drastically improved their writing compared to those who did not. Finally, we discuss future research in self-monitoring of writing.

Keywords. Self-monitoring, Self-evaluation, Peer evaluation, Peer review, Writing, CSCL, SWoRD
Self-Monitoring Support in Learning to Write

While writing is considered as one of the most important skills that students are expected to master for professional as well as academic success, writing well is a fundamental skill that many students lack across a wide range of ages in the U.S. (National Commission on Writing, 2003) and also very likely in other countries. Accordingly, the U.S. National Commission on Writing, representing more than 4,300 schools and colleges in the U.S., declared great urgency to increase the emphasis on writing at all levels of education.

To write well, students are expected to have self-regulation skills (Bereiter & Scardamalia, 1987; Graham & Harris, 2000; Hayes & Flower, 1980; Zimmerman & Kitsantas, 1999). For example, experienced writers successfully set goals, self-monitor, and reflect on their learning process through iterative writing and rewriting. Among the various self-regulation skills, self-monitoring is known to be particularly critical in learning to write (Cresswell, 2000; Xing, 2004). Self-monitoring involves accessing information about ongoing cognitive or affective processes, states, or memory retrieval and evaluating that information on either implicit or explicit criteria (Reder & Schunn, 1996; Shimamura, 1996). In the context of writing, self-monitoring involves evaluating information about what is currently being written or what has been written and noting some compatibility or incompatibility with a mental representation of what the written text should be (Hacker, 1994). The greater the incompatibility, the more likely the writer will engage in revision. The greater the compatibility, the less likely the writer will need to revise.
This study examines the critical role of self-monitoring in writing and whether writing is improved by providing self-monitoring support. We investigated two research questions: (a) Do students improve their self-monitoring of writing over time, and (b) are improvements in self-monitoring related to improvements in writing quality? We examined these questions by comparing students’ self-assessments of their writing with the assessments from a peer group. We used a computer supported writing environment called Scaffolded Writing and Revision in the Disciplines (SWoRD). Because so little empirical research on computerized writing environments has been conducted (Cresswell, 2000), we believed that this component of our research would be a valuable contribution to the literature.

Self-Monitoring during Writing

Existing research agrees upon the critical role of self-monitoring during writing (Graham, Harris, & Mason, 2005; Winne & Hadwin, 1998). While experienced writers have good self-monitoring skills, inexperienced writers do not. For example, before writing, experienced writers monitor various textual components for better communication with readers. Experienced writers are found to be aware of their writing limitations, necessary processes, and how audience would respond to their writing (Zimmerman & Risemberg, 1997). By contrast, inexperienced writers have difficulties self-monitoring their writing. For example, they are often unsuccessful in detecting problems in texts and also unsuccessful in fixing problems even if they detect them (Bracewell, Scardamalia, & Bereiter, 1978; Hacker, Plumb, Butterfield, Quathamer, & Heineken, 1994). Inexperienced writers tend to monitor local problems such as words or grammatical errors, unlike experienced writers who monitor global or structural levels.
(Butterfield, Hacker, & Plumb, 1994; Faigley & Witte, 1983; Sommers, 1980). In addition, not surprisingly, inexperienced writers are rarely aware of readers’ perspectives (Gearhart & Wolf, 1994). Therefore, for inexperienced writers to acquire greater writing skill, they need support to more fully develop self-monitoring of their writing processes.

Self-monitoring can be an implicit or explicit cognitive process (Reder & Schunn, 1996; Simon, 1996) the point of which is to self-assess the compatibility of some information with a mental representation of that information (Schacter, Bowers, & Booker, 1989). There are numerous examples of implicit self-monitoring that occur during writing (e.g., monitoring of language production or graphomotor production during typing or handwriting). However, the focus of the present study was on explicit monitoring, specifically, the degree to which students’ self-assessments of their writing on specific criteria conform to the assessments from a peer group of readers on those same criteria. Thus, we defined self-monitoring as a self-assessment skill in which writers perceive their writing from the perspective of readers (Butler & Winne, 1995; McCaslin & Hickey, 2001).

Although participating in self- and peer-evaluation of writing is expected to improve self-monitoring skills and writing quality (Cho, Schunn, & Carney, 2006; Herrington & Gadman, 1991; Marcoulides & Simkin, 1991; Shaw, 2002), a reasonable concern is that students may not be accurate self-monitors of their writing. Indeed, Hayes and Flower (1980) found that inaccurate monitoring might undermine its positive role in writing improvement. When student writers overestimate or underestimate their writing quality, their inaccurate monitoring may hinder them from setting realistic goals and from using appropriate writing strategies. In their coregulation model, McCaslin and Hickey (2001)
emphasized the importance of consistency of self-monitoring results between self- and peer-evaluations. According to them, if students over- or under-estimate their learning in comparison to how others estimate their learning, it can lead to inappropriate use of strategies and to less skillful self-regulation. Thus, self-monitoring of writing processes may be fine-tuned with explicit comparisons between self- and peer evaluations.

**Self-Monitoring Support in SWoRD**

While self-monitoring is a critical component for learning to write (Cresswell, 2000; Gearhart & Wolf, 1994; Graham & Harris, 1993; Xing, 2004), providing support to develop self-monitoring is not straightforward in the reciprocal peer reviewing of writing. To provide this support, Cho and Schunn (2007) developed SWoRD, which is a web-based, hybrid intelligent system for writing delivery and assessment that students and teachers use for peer review. Since its creation in fall 2002, about 4000 students at 14 universities in 70 courses have used SWoRD across the disciplines of physics, psychology, education, history, law, information technology, composition, and English education. SWoRD is equipped with two types of self-monitoring supports: One is for students to compare their self-assessments and peer assessments on their own writing, and the other is for students to compare their assessments and peer assessments on others’ writing (see Figure 1).

----------

Insert Figure 1 about here.

----------

**Comparing Self-Evaluation and Peer-Evaluations on Their Own Writing**

Unlike traditional classrooms where instructors are the most typical and single source of evaluation, with reciprocal peer reviewing, student reviewers are often asked to review
draft papers from multiple peers (Cho & Schunn, 2005; Cho, Schunn, & Wilson, 2006). Therefore, authors are provided with diverse evaluations on their own writing. In addition, SWoRD requires that authors self-evaluate their own writing using the same criteria that reviewers use. Therefore, authors’ self-evaluations may help them to develop more explicit awareness of their writing quality or problems as seen from multiple readers’ perspectives.

Students also are allowed to compare their self-evaluations on their own writing with peer evaluations. The discrepancies between the two evaluations can trigger authors to act upon the gaps to improve their writing. We should note that in the present study SWoRD did not explicitly ask students to make comparisons between self-evaluation and peer evaluation on their own writing. However, there were ample opportunities for students to make these comparisons.

Comparing Self-Evaluation and Peer-Evaluations on Others’ Writing

The other type of monitoring support provided by SWoRD was designed to allow each reviewer to compare their own evaluations with other evaluations on peer papers. By evaluating peer writing and seeing the evaluations of that writing from other peers, students can better understand the writing assignment, evaluation rubrics, or what to do or not do in writing (Butler & Winne, 1995; Cho & Schunn, 2007). This active comparison of self and others’ evaluations may enable students as reviewers to view writing from various reader perspectives, and further, to enable students as writers to improve their own writing.

Figure 2 shows an example of the reviewer support interface activating self-monitoring for a reviewer. The interface visualizes the extent to which a reviewer’s
evaluation is consistent with that of others who reviewed the same papers. The pattern in Figure 2 shows that the reviewer’s grades are consistent with those of others, while there is a visually significant difference with reviewer 400lb Gorilla. Pseudonyms such as 400lb Gorilla or River are used to keep students from identifying reviewers. Also, when a reviewer clicks on the author name, the reviewer can then read both her own review and others’ reviews on the same writing.

In sum, the goal of this study was to examine the role of self-monitoring support in improving writing skills. More specifically, we investigated (a) how self-monitoring of writing would change over time with exposure to one’s own evaluations of writing and to others’ evaluations of writing, and (b) whether self-monitoring is related to writing improvement. Unlike past self-regulation research, which has generally been conducted with a small number of elementary or adolescent students (e.g., Graham & Harris, 1993; McCaslin & Hickey, 2001; Vygotsky, 1978), we used a large number of participants in three U.S. universities.

Method

Participants

The participants in the study were 601 students from three research universities in the U.S. across 16 courses representing various disciplines (e.g., cognitive psychology, cognitive science, physics, and health psychology). Three of the courses were graduate level, 13 were undergraduate level, and four among the 13 undergraduate courses were for non-majors (see Table 1). The instructors of the courses used the SWoRD system to
support their writing assignments. The students participated in this research as part of their regular course activities. Although the writing and reviewing assignments together accounted for approximately 40% of the final course grade in each course, the actual grade proportion did vary depending on the course.

---

Insert Table 1 about here.

---

**Writing Task**

As one would expect across courses from many different disciplines, the writing task assigned to students varied across the courses. The required length of the assigned papers varied from shorter (5-to-8 pages) to longer papers (10-to-15 pages). Writing genres included (a) a research paper, (b) an application to real life of research findings, (c) technical writing, or (d) a research proposal. Although the genre and length did differ across courses, the self- and peer-review process was identical for all courses. We believe that these differences in writing tasks helped to increase the generalizability of the study.

**Measures**

*Dimensions for evaluating writing.* Authors and reviewers used the same three dimensions for evaluating papers: prose flow, argumentation, and insight. Prose flow dealt with how smoothly the writing flowed so that the reader could easily understand and could easily transition from one point to the next; argumentation dealt with the logic of the arguments being made in the paper; and insight dealt with whether the author had introduced any new information in the paper beyond what was provided in class. Papers were evaluated on these three dimensions using a 7-point rating scale (1 = disastrous to 7 = excellent).
Writing quality. The quality of a student’s paper was measured by calculating the average of the peer reviewers’ evaluations on each of the three dimensions of writing. Improvement in writing quality was calculated by taking the difference between writing quality on the first and second drafts.

Self-monitoring. Self-monitoring (SM) was defined as an absolute difference between self-evaluation (SE) and other evaluation (OE) on a student’s writing:

$$SM = |SE - OE|$$

Differences closer to zero indicated that a student was better at self-monitoring. OE was an average of other evaluations, and SE was an average self-evaluation. OE and SE were calculated using the first writing and revised writing.

Self-monitoring development. Self-monitoring development (SMD) was defined as a unidirectional change from the first self-monitoring (SMt1) to the second self-monitoring (SMt2):

$$SMD = SMt1 - SMt2$$

A positive SMD indicates that self-monitoring was developed because positive values mean SMt2 is smaller than SMt1. This means the gap between self- and peer-evaluation is reduced over time. By contrast, if SMD is equal to or less than zero, SMD was defined as not-yet-developed. This means the gap between self- and peer-evaluation is either the same or is not reduced over time.

Procedure

The experiment followed the built-in processes in SWoRD. SWoRD is a Web-based application (http://www.missouri.edu/~chokw/SWoRD.html) that (a) helps manage the
distribution of papers to reviewers and reviews back to authors and (b) includes evaluation mechanisms that support students to self-monitor their writing through self-evaluation and peer evaluation.

As shown in Figure 3, in step one, students create an account in the system and specify a pseudonym. This pseudonym is used to reduce any status biases that may occur in peer review. Reviewers are identified to authors only by number (e.g., reviewer #1, reviewer #2, etc) to ensure there is no retribution between particular authors and reviewers. At this time, instructors set writing and reviewing assignments, due dates, and assignment policies.

In step two, which was optional, students practiced reviewing sample papers using the three dimensions for writing quality (i.e., prose flow, argumentation, insight). They then received feedback from SWoRD based on past expert review and peer reviews on the same papers.

In step three, student authors logged into the SWoRD system to upload their draft papers before the first draft deadline. When submitting their drafts, students also self-evaluated their writing quality based on the three dimensions. Once the submission deadline passed, each author’s draft paper was assigned to $n$ peers, where $n$ was pre-specified by the instructor (usually 5 or 6). We used a moving window algorithm so that no two drafts were assigned to the same set of $n$ peers. Also, student reviewers were selected based on the probability of each student completing reviews and of being a fair reviewer to maximize the chance of writers’ receiving constructive peer reviews.
In step four, peer reviewers logged into the system and downloaded the assigned papers and a copy of instructions on how to evaluate the papers on the same three dimensions. For each evaluation dimension for each paper, the reviewer entered written feedback on what problems were found in the paper and provided specific suggestions for improving the paper. Reviewers were given specific suggestions for how to structure the written comment. For example, for the prose flow dimension, reviewers were asked to briefly summarize the main argument of the paper so authors can see how well their main points came across. The reviewer also gave a numerical rating on a 1-7 scale. Each rating point was given a clear anchor—what features a paper should have to deserve the given rating in order to maximize consistency across reviewers. Once the reviews are uploaded into SWoRD, the system evaluates each reviewer’s evaluation in terms of three measures: problems in relative ordering of paper quality, systemically high or low evaluation, and systematic problems in how broadly or narrowly evaluations are made. The goal of the grades is to force accountability on the peer- grading task and to encourage reviewers to consider a broader audience than just themselves.

In step five, student authors accessed the full set of comments on their paper, the system’s assessment of each reviewer’s consistency (marked with stars), their overall writing grade so far in relation to the class mean, the system’s assessment of their own reviewing consistency, and their overall reviewing grade so far relative to the class mean. In light of all this feedback, students then revised their papers and uploaded their final drafts, which were then distributed back to the same peer reviewers as in the first round of reviewing.
In step six, once the draft paper was submitted, each author was asked to rate the helpfulness of each review they received. They used a 7-point helpfulness scale, from not helpful at all (1) to very helpful (7). These ratings constitute the other half of the reviewer’s reviewing grade and serve to encourage reviewers to take the written review task seriously.

In step seven, reviewers downloaded the final drafts assigned to them and reviewed the draft papers for the final time, using the same rating rubric that was used as for the first draft. Once again, the reviewers were automatically assessed by the SWoRD system for: (a) the accuracy of their numerical ratings, and (b) the authors’ helpfulness of their written comments. Note that reviewers also had access to the ratings and comments given by other reviewers for their set of papers. In this way, they could learn more about other perspectives and reactions to different prose through the reactions other students had to the same papers.

In step eight, students were provided with the full set of comments on their draft paper, the system’s assessment of each reviewer’s consistency, their overall writing grade so far in relation to the class mean, the system’s assessment of their own reviewing consistency, and their overall reviewing grade so far relative to the class mean. Students were again asked to grade the helpfulness of the final draft comments. SWoRD automatically placed equal weight on first and final draft activities and equally weights reviewing rating consistency and comment helpfulness.
Results

*Did SM change over time?*

We first examined if students improved their self-monitoring of writing over time while they self-evaluated and received feedback from peers. For the 601 students, the mean SMt1 was 3.04 (SD = 2.6) and SMt2 was 3.32 (SD = 3.1). Among 601 students, 287 students developed their SM skills over time [Kwangsu: could you provide statistics on those who improved along with an effect size?] and 314 students did not develop their SM skills. In general, SM significantly decreased from SMt1 to SMt2, t (600) = -2.23, p < .05.

Paired t-tests were conducted to investigate SM change over time in each course of the 16 courses. Among the 16 courses, significant changes appeared only in three courses (G03: t(29) = 2.67, p < .05; U02: t(16) = -3.47, p < .05; U05: t(69) = 2.06, p < .05).

[Kwangsu: could you provide means, SDs, and effect sizes on these three?]

*Was SM change linearly associated with writing improvement?*

To address whether improvements in self-monitoring were related to improvements in writing quality, we first calculated the SMD for all 16 courses and then calculated a Pearson correlation between the SMD and writing quality improvement. Remember that a positive SMD indicates that self-monitoring was developed and that greater SMD values indicate greater improvements in self-monitoring from time one to time two. The correlation was statistically significant, r (600) = .66, p < .001, showing that as self-monitoring of writing improved so did writing quality.

We then calculated the Pearson correlations for each of the 16 courses. Figure 5 shows individual correlations between SMD and the writing quality improvement with
95% confidence intervals bars. The Pearson correlations were significant for two of the three graduate courses and for 11 of the 13 undergraduate courses, $p < .05$. Also, the ranges of correlations were diverse from -.18 to .82. The average Pearson correlation over the three graduate courses was .58 and was .55 for the 13 undergraduate courses.

Discussion

In this study, we examined whether students improve their self-monitoring (SM) of writing over time and whether improvements in self-monitoring are related to improvements in writing quality. We defined SM as a self-assessment skill in which writers perceive their writing from the perspective of readers (Butler & Winne, 1995; McCaslin & Hickey, 2001), and we measured it by comparing students’ self-assessments of their writing on specific criteria to the assessments from a peer group of readers on those same criteria. Students’ self-assessments and peer assessments were recorded, managed, and measured through the use of SWoRD, a computer supported writing environment.

The results showed that SM skills do not change easily: Only about one-half of the participants improved their SM skills over time. For these participants, their self-evaluations of writing quality over time moved closer to their peers’ evaluations. If we assume that the peer evaluations, which were based on the averaged evaluations from 5 or 6 peers, represent a more accurate appraisal of writing quality—an assumption that is not difficult to accept—then we can conclude that many of our writers become more skilled at seeing their writing through the perspective of their readers (i.e., they became more accurate at SM).
More research is necessary to investigate why some students did not change their SM skills. Peer review interventions have not all produced positive results. Students can have considerable difficulty understanding the rationale behind peer feedback, judging the relative worth of peer evaluations, or knowing how the feedback, if implemented, will improve their writing (Hull, 1987). Students can receive feedback from peers, but without deeper understanding of what is contributing to the feedback or knowing whether the suggested changes are necessary, students may be led to making revisions to their writing that actually decrease writing quality. Therefore, knowing why nearly one-half of our participants did not improve their SM in response to peer feedback is critical.

In addition, the research results showed that for students in higher education, SM is an important correlate to writing quality. Consistent with self-regulated learning theories, this study empirically showed that students who showed strong gains in SMD—an essential component of self-regulated learning—also showed strong gains in writing quality (Hayes & Flower, 1980; Winne & Hadwin, 1998). This result seems consistent with Graham and Harris’s self-regulated strategy development (SRSD) model in that self-regulated writing positively influences writing performance (Graham & Harris, 1993). Moreover, Graham, Harris, and Mason (2005) found that students who received SRSD training significantly improved their writing in their posttest as well as outperformed those who did not receive SRSD training in terms of writing length, story element, persuasive elements, and quality. Because the present study was a large-scale field study, it has stronger generalizability in comparison to past studies of self-regulated writing and adds considerable credibility to arguments for the need of self-regulation to improve performance.
Although this study endorses the role of SMD in writing improvement, careful interpretations are necessary. First, we did not provide explicit instructions on SM to the students in SWoRD. This approach is somewhat different from the previous research on self-regulated learning that emphasizes explicit training of self-regulated learning strategies (Graham & Harris, 1993). For example, Zimmerman, Bonner, and Kovach (1996) suggest that students be explicitly instructed on the benefits of using self-regulated learning strategies before and also while performing tasks. Students may better benefit from SM when they are aware of the role of SM in learning to write. Relatedly, explicit comparisons between self- and peer evaluation on self-writing that were not provided for students would help them reflect upon the gaps between the evaluations.

Second, although students were provided with the self-monitoring interfaces, how the students used the interfaces was not examined. How students used information would allow researchers to look into self-monitoring behaviors at a more detailed level. The results would reveal how students use the system for their self-monitoring. This information also would be important to design self-monitoring support because students in the computer environment are less likely to activate self-regulated learning strategies than those in the human-agent environment (Azevedo, 2005).

Finally, qualitative analysis on reviews might reveal how feedback quality influences self-monitoring in learning to write. Students who receive quality feedback from peers may be more apt to accept suggestions and self-regulate to bridge the gap between self-evaluation and peer evaluation. By contrast, students who receive low quality feedback from peers may ignore suggestions and make less effort to bridge the gap between self- and other evaluation.
Acknowledgement

This research was supported by the grants from the Andrew Mellon Foundation to the first author. We thank Carla Bates for her comments on an earlier draft.


Shaw, V. N. (2002). Peer review as a motivating device in the training of writing skills for college students. *Journal of College Reading and Learning, 33*, 68-76.


Table 1

*Participant Information*

<table>
<thead>
<tr>
<th>Course ID</th>
<th>Discipline</th>
<th>Year</th>
<th>University</th>
<th>No. of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>G01</td>
<td>Health Psychology</td>
<td>2005</td>
<td>A</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Instructional</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G02</td>
<td>Technology</td>
<td>2004</td>
<td>C</td>
<td>12</td>
</tr>
<tr>
<td>G03</td>
<td>Biomedical Informatics</td>
<td>2005</td>
<td>B</td>
<td>14</td>
</tr>
<tr>
<td>U01</td>
<td>Cognitive Science</td>
<td>2004</td>
<td>B</td>
<td>20</td>
</tr>
<tr>
<td>U02</td>
<td>Cognitive Science</td>
<td>2005</td>
<td>B</td>
<td>17</td>
</tr>
<tr>
<td>U03</td>
<td>Cognitive Psychology*</td>
<td>2005</td>
<td>B</td>
<td>67</td>
</tr>
<tr>
<td>U04</td>
<td>Cognitive Psychology*</td>
<td>2005</td>
<td>B</td>
<td>78</td>
</tr>
<tr>
<td>U05</td>
<td>Cognitive Psychology*</td>
<td>2004</td>
<td>B</td>
<td>70</td>
</tr>
<tr>
<td>U06</td>
<td>Cognitive Psychology*</td>
<td>2005</td>
<td>B</td>
<td>25</td>
</tr>
<tr>
<td>U07</td>
<td>Physics</td>
<td>2005</td>
<td>B</td>
<td>81</td>
</tr>
<tr>
<td>U08</td>
<td>Physics</td>
<td>2005</td>
<td>B</td>
<td>26</td>
</tr>
<tr>
<td>U09</td>
<td>Writing</td>
<td>2005</td>
<td>B</td>
<td>10</td>
</tr>
<tr>
<td>U10</td>
<td>Writing</td>
<td>2005</td>
<td>B</td>
<td>19</td>
</tr>
<tr>
<td>U11</td>
<td>History</td>
<td>2005</td>
<td>B</td>
<td>97</td>
</tr>
<tr>
<td>U12</td>
<td>Leisure Behavior</td>
<td>2005</td>
<td>C</td>
<td>19</td>
</tr>
<tr>
<td>U13</td>
<td>Leisure Behavior</td>
<td>2005</td>
<td>C</td>
<td>16</td>
</tr>
</tbody>
</table>

*Note.*  G = graduate; U = undergraduate

* Courses for non-majors.
Figure Captions

Figure 1. Self-monitoring support in SWoRD

Figure 2. An example interface in SWoRD

Figure 3. Major SWoRD processes

Figure 4. Pearson correlation between SMD and writing quality improvement
Figure 1. Self-monitoring support in SWoRD
Figure 2. An example interface in SWoRD
Figure 3. Major SWoRD Processes
Figure 4. Pearson Correlation between SMD and Writing Quality Improvement

G: Graduate course; U = Undergraduate course, *: p < .05, n: number of students