

Learning How to Write through Encouraging Metacognitive Monitoring: The Effect of Evaluating Essays written by Others

Miwa Inuzuka (miwainzk@aol.com)¹

Graduate School of Education, University of Tokyo, 7-3-1 Hongo Bunkyo-ku
Tokyo 113-0033, Japan

Abstract

This study investigated how facilitating metacognitive monitoring could improve the writing performance of students. To this end, an intervention called Evaluating Activity (EA) was developed, and its effect was investigated. In Study 1, EA was implemented for 30 eighth-grade students. In EA, students are told to read and rate sample essays. After evaluating an essay, they are required to comment on the strengths and weaknesses of the text. The analysis of the participants' comments suggested that a variety of viewpoints was crucial for improving their writing performance. In Study 2, 91 eighth-graders participated. To encourage the participants to consider a variety of viewpoints, collaborative work was introduced. The purpose of Study 2 was to investigate differences between three conditions of EA: individual (I), collaborative (C), and collaboration following individual work (I-C). The results indicated that the writing performance of the participants improved significantly under Condition I-C. In addition, the advantage of a variety of viewpoints increased only in Condition I-C.

Keywords: Metacognitive monitoring; writing skills; Evaluating Activity; collaboration.

Introduction

How do people learn to write, and how can we improve students' writing skills? Based on the process model of writing (Hayes & Flower, 1980), numerous interventions have been developed and studied. In many of these studies, the writing process was divided into components, and strategies to manage those components were developed (Bereiter, Scardamalia, & Steinbach, 1984; Galbraith & Rijlaarsdam, 1999 as a review). Most of these interventions were designed to teach participants strategies for planning (Kellogg, 1988) or revising the essay (Zimmerman & Kitsantas, 1999). These studies showed that such strategies enhanced students' writing skills (Sawyer, Graham, & Harris, 1992). In addition, the process model of writing itself suggests that a writer's metacognition plays an important role throughout the writing process (Flower & Hayes, 1980).

Metacognition has been defined as having two different components: metacognitive activity and metacognitive knowledge. In this study, we concentrate on metacognitive activity in the writing process. Metacognitive activity is composed of two different processes: monitoring and control. Monitoring is the process through which people judges discrepancies between the present state and the goal state

(Nelson & Narens, 1990). On the other hand, control process involves with the implementation of the strategies.

Regarding the relationship between writing process and the two aspects of metacognitive activity, several studies have pointed to the role of metacognitive control. For example, Breetvelt, Van Den Bergh, & Rijlaarsdam (1994) revealed the metacognitive control affects writing process by indicating that the time at which a writer engaged in a particular process, and what they did, were related to the quality of writing. In order to employ appropriate strategies at appropriate timing, it is necessary for the writer to correctly recognize his/her goal and present state. Surprisingly, however, only few studies have investigated the effect of metacognitive monitoring on essay writing. Therefore, in this study, we investigated how we could encourage metacognitive monitoring, and its effect on writing performance.

There were two reasons behind our focus on metacognitive monitoring. (1) Since previous research on learning strategies had suggested that monitoring ability affects the choice of strategy and performance (Thiede & Dunlosky, 1999; Thiede, Anderson, & Therriault, 2003), it seemed important to investigate the role of monitoring in relation to the writing process. (2) Although researchers have already advocated the importance of monitoring, students do not get enough instruction in schools in this regard.

To encourage students to enhance their monitoring skills, we developed an intervention called Evaluating Activity (EA), which we introduce here. In this intervention, students first read essays with the explicit aim of learning the principles that are essential for writing a good essay. Participants read essays that another person has written, and rate them on a 5-point scale (1-very poor to 5-excellent). Then, the participants comment on why they rated the essay as they did, and suggest ways to improve the essay.

EA was designed so that students can adopt an objective point of view, by separating the judging activity from the writing process. It is difficult to monitor one's own essay while writing it, because the writer needs to involve two different cognitive levels when writing. Separating the two levels makes it easier for the participants to conduct monitoring activity. We included a comment-making activity, so that the participants could look back on their activity. This was aimed at helping the participants to clarify their monitoring activity. In addition, the comments allowed us to perceive their views on their monitoring. In addition, we

¹ The author is now at Japan Society for the Promotion of Science / Saitama Institute of Technology. 1690 Fusaiji Oaza Okabe-cho Osato-gun Saitama 369-0293, Japan

provided participants with essays written by strangers, in order to help the participants examine essays critically. Furthermore, we included a suggestion making activity, as a bridge between the monitoring and the control activity.

Since the activity is based on the students' evaluation, and no other training takes place, the intervention introduced here is called Evaluating Activity (EA).

In subsequent studies, we investigated whether EA improves students' writing performance. In order to investigate how the participants evaluated or monitored the essays, we analyzed the participants' comments.

Study 1

Study 1 investigated the effectiveness of the evaluating activities.

Method

Participants: Thirty eighth-grade students participated in the study. All the participants were public school students, who participated in the study during classes for a summer course held at the University of Tokyo, in which they had independently enrolled.

Materials All the sample essays presented to the participants were chosen from a reference book on essay writing. Students had originally written the essays; therefore, not written skillfully. Some of the essays contained grammatical mistakes and were not sufficiently organized. These essays were chosen for the following reasons: (1) the participants would regard the presented essays as imperfect models that could be improved upon, and (2) as the author was anonymous, it reassured the students that their own writing would not be subjected to the EA.

Procedures Before the EA took place, the participants were first instructed to write an essay on one of two topics: "In your opinion, which one of the following offers the better mode of transportation, cars or trains?" or "In your opinion, which of the following offers the better way to obtain information, TV or newspapers?" The essays were to be analyzed as the pre-test. The topic was counterbalanced among participants. Subsequently, the EA took place. It was conducted as a four-day session. In each session, the students were given a sample essay and instructed to evaluate it. The same set of essays was given to all of the participants. Approximately 25 minutes were allotted for the entire activity. A worksheet was prepared for the EA. Using the worksheets, the participants (1) rated the essay, (2) commented on the essay and described its strengths and weaknesses, and (3) suggested ideas for improvement. At the beginning and end of each class, the instructor reminded the participants of the aim of the activity: The activity was conducted so that the participants would learn important viewpoints involved in monitoring on writing, as well as the ability to apply these viewpoints in monitoring when writing their own essay. However, the instructor did not provide them with feedback on their comments. At the end of the four-day session, the

participants were instructed to write an essay on the other pre-test essay topic.

Results and Discussion

Performance in the Pre-test and Post-test The data for 28 participants, who participated in all of the sessions, were analyzed. All the essays were typed, and misspelling and grammatical mistakes were corrected. Then, the author and three graduate students rated the essays independently. All the raters rated the essays on a 5-point Likert-type scale from 1 (not at all) to 5 (highly). The essay was rated on the following: (1) readability, (2) persuasiveness, and (3) coherence. The mean scores for each were calculated across raters. Then, the scores were summed, giving a maximum score of 15 for an essay.

The mean pre-test score was 10.17 (SD=1.95) and the mean post-test score was 10.06 (SD=1.31). The scores were analyzed using within-subject analysis of variance and did not differ significantly ($F < 1$).

Although there was no significant difference between the pre-test and post-test scores, we found that half of the participants ($N=14$) had improved their writing performance, while the performance of the other half had deteriorated. Furthermore, a change exceeding one standard deviation was observed for 11 participants.

Participants' Comments Then, we investigated the viewpoints of the participants. To this end, the comments made during the last session were analyzed to discover differences between the improved and deteriorated groups. The comments of the participants whose performance had changed more than one SD were selected. An examination of each participant's performance change revealed that five participants improved by more than 1 SD, while six deteriorated by more than 1 SD. The author and one graduate student, who did not know the condition of the participants, analyzed the comments. First, we chose keywords in the comments (e.g., "The essay was hard to understand because he changed the topic in the middle, but it was good that he

Table 1. Categories and examples of the participants' comments.

Categories	Examples of comments
Coherence	(The author) changed the topic in the middle (of the essay)
Construction of the essay	(The author) made paragraphs appropriately
Conclusion	(The author) described his or her own opinions
Reader-friendliness	The essay was difficult to understand
Language usage	(The author) mixed formal and casual lines in the same essay
Content of the essay	I do not think the author's claim is correct.

Table 2. Number of participants' comments on the example essay. (% of all comments)

	Improved	Deteriorated
Coherence	2 (6 %)	1 (2 %)
Construction of the essay	6 (18 %)	3 (7 %)
Conclusion	6 (18%)	7 (17%)
Reader-friendliness	2 (6 %)	0 (0 %)
Linguistic usage	9 (26 %)	9 (21 %)
Content of the essay	8 (24 %)	22 (52 %)

described his own opinions.”). Then, we divided the comments that had more than two keywords so that each comment contained one keyword (e.g., “The essay was hard to understand”, “(The author) changed the topic in the middle (of the essay)”, and “He (the author) described his own opinions”). According to the keywords, the participants' comments were divided into six categories: (1) coherence, (2) construction of the essay, (3) drawing his/her own conclusion, (4) how easy the essay was to understand, (5) grammatical mistakes and phrasing, and (6) comments on the essay content (Table 1). All of these comments were categorized separately by the author and the graduate student, who did not know the participants' condition. Then, the comments were judged by another graduate student, who did not know the participants' condition. Table 1 suggested that the participants whose performance was improved made more various comments. The comparison of the mean number of comment categories between two groups confirmed the impression revealing that the participants in improved group made larger number of comment categories than the participants in deteriorated group did ($p < .10$).

The results showed that the effect of the EA did not reach significance, and that its impact differed widely across participants. These results suggested that the process that was followed during the activity differed across participants. The investigation of the participants' comments suggested that the participants with more varied viewpoints improved their own writing performance. Conversely, the writing performance of the participants who had biased viewpoints, and who focused on specific content, deteriorated. These results suggested that the EA was effective when participants generate diverse viewpoints. Therefore, a variety of viewpoints of monitoring would be a key to improving their writing performance. The results implied that in order to improve the participants' performance, it is important to produce a situation that encourages the participants to learn diverse viewpoints.

Study 2

From the results of Study 1, we hypothesized that if the participants succeeded in perceiving a variety of viewpoints during metacognitive monitoring, then their writing performance would improve. Study 2 was aimed at encouraging participants to achieve more diversity in their viewpoints of monitoring, and investigating its effects.

To this end, we revised the original EA and added time for listening to other peoples' ideas, with the instruction that participants should learn from other people's ideas. We

expected that this change would help the participants to develop more variety in their viewpoints of monitoring. In addition, we introduced a collaborative activity to the EA. Cognitive science studies on collaborative learning have shown that collaboration facilitates problem-solving (Miyake, 1986; Okada & Simon, 1997; Kiyokawa, Okada, & Ueda, 2002), learning the correct concepts (Schwarz, Neuman, & Bieauner, 2000), and achieving diversity with respect to viewpoints (Shirouzu, Miyake, & Masukawa, 2002). Based on these previous researches, we hypothesized that discussing the essay would lead participants to think collaboratively and thus facilitate the achievement of diverse viewpoints of metacognitive monitoring.

The studies mentioned above have reported the positive effects of collaboration on problem-solving; nevertheless, some studies have reported a negative effect of working in groups, namely groupthink (Janis, 1972; Turner & Pratkanis, 1998). Groupthink is a phenomenon that occurs in group decision making. When a group of people are required making some decision, some negative effects of collaboration emerge: the group examines few alternatives, and become uncritical of other's ideas. Participants might regard group activity in EA as decision making task, because it requires the participants to state the group's ideas. If this is the case, the collaborative activity in EA ends up with limited viewpoints. To encourage participants' viewpoints during monitoring to become more diverse through discussion, it is necessary to avoid the negative effects of groupthink.

For this purpose, a condition that required the participants to work as individuals, before participating in the group discussion, was introduced. Having prepared the presentation of one's own ideas, participants were able to examine each member's ideas. For this reason, the individual work would help the participants discuss broader ideas, and thereby, the possibility of opposite opinion would increase. Therefore, we hypothesized that negative effects of groupthink would not occur with an individual work condition, and that the students would succeed in achieving the various viewpoints. Conversely, when lacking an opportunity to prepare their own ideas, the participants' ideas would be limited. In this case, the viewpoints of participants in the collaboration-only condition would be likely to become more and more biased, and they would fail to achieve a variety of viewpoints.

Method

Participants The participants in this study were 91 eighth-grade students. All the participants were students at public schools. They participated in the study during classes for a summer course, held at the University of Tokyo, in which they had enrolled independently. The participants were randomly assigned to three experimental conditions described below.

Materials All the sample essays were the same as those used in Study 1.

Procedures The procedure followed was the same as that for Study 1, except that the participants were divided into

three conditions. In the collaboration condition (Condition C), the participants worked in groups of three. They rated and described the sample essay's strengths and weaknesses, and suggested ideas for improving it through discussion. In the individual condition (Condition I), the participants engaged in the same EA as in Study 1, except for the instruction that directed them to learn from others' ideas. In the individual-collaboration condition (Condition I-C), the participants first rated the sample essay and described its strengths and weaknesses. Then, they formed a group of three and discussed each group member's ideas, and suggested possibilities for improvement. Although the activity styles differed, the amount of time they spent on the EA was approximately the same (25 min.). Another change from Study 1 was in the pre-test/post-test writing topics ("In your opinion, which of the following is the better mode of transportation, car or train?" or "Some people argue that 'language pollution' is a serious problem among young people. What do you think about this claim?"). The topics for the pre-test/post-test were changed because the topics used in Study 1 were too similar. The writing topics were counterbalanced among participants.

Results and Discussion

In the following, the significance level was $p < .05$ unless otherwise mentioned. The data for the participants who participated in all of the sessions were analyzed.

Performance in the pre-test and post-test: The scoring procedures were essentially the same as in Study 1. Only the items rated were changed, to make the rating more sensitive and reliable. In addition to rating the impression of persuasiveness, we rated the strength of the argument, the extent to which an appropriate basis for the conclusion had been provided, and the clarity of the conclusion. Since the above-mentioned items relate to the coherence of the essay, we did not rate coherence this time. Instead, we rated the structure of the essays and whether the essay had appropriate paragraphs. Consequently, the pre-test and post-test essays were scored on the following factors: (1) readability, (2)

persuasiveness, (3) strong argument, (4) appropriate basis for conclusion, (5) strong conclusion, and (6) appropriate paragraphs.

Since the scores were correlated with each other ($\alpha = .86$), a total of the six scores was used in the subsequent analysis. For the participants in Conditions C and I-C, the average scores of each group were analyzed. Therefore, the number of data in Conditions C, I-C, and I were 7, 8, and 18, respectively. The mean scores for each condition were as follows: the respective mean pre-test and post-test scores were 21.58 (SD=6.04) and 19.37 (SD=6.14) in Condition I, 21.61 (SD=3.98) and 21.00 (SD=4.94) in Condition C, and 21.37 (SD=4.95) and 24.98 (SD=4.32) in Condition I-C (Figure 1). A General Linear Model (GLM) procedure was conducted, with the condition and pre-test scores as the independent values, and the post-test scores as the dependent value. The results of the GLM revealed that the main effect of pre-test and condition was significant ($F(1,28)=7.58$, $F(2,28)=3.79$, respectively). The other effects did not reach significance. The subsequent multiple comparison revealed that the post-test performance of the participants in Condition I-C was significantly better than that of the participants in Condition I.

Table 3 shows the change in the participants' scores. The participants/groups whose scores improved by more than 1 SD were considered "Improved", while the participants/groups whose scores deteriorated by more than 1 SD were considered "Deteriorated", and the other participants/groups were considered "Unchanged". The scores of the participants in Conditions C and I changed, just like they did for the participants in Study 1. The results of Condition I replicated Study 1. In addition, the results of Condition C showed a tendency similar to that of Study 1. However, a different tendency was found in the results of Condition I-C, in which most of the groups showed improved writing performance and only one group in Condition I-C showed deterioration.

Participants' Comments The participants' comments were divided into six categories, as in Study 1. The coding procedures were the same as in Study 1. Note that the comments mentioned below were made after the discussion in Conditions C and I-C. By contrast, those of Condition I were made before listening to others' ideas.

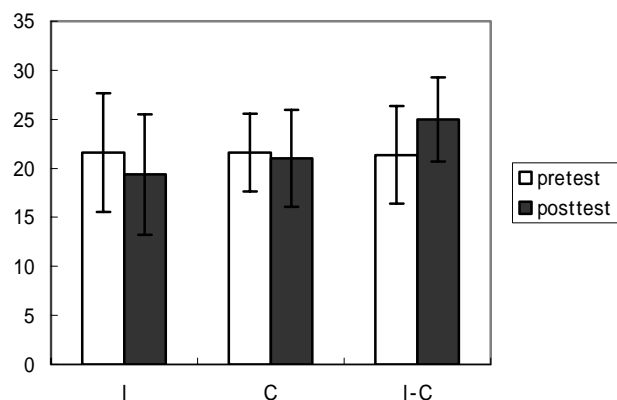


Figure 1. Pre-test and Post-test performance in each condition. I: individual condition, C: collaboration condition, and I-C: collaboration-individual condition.

Table 3. Number of participants/groups distributed according to the change in scores

Condition	Improved	Unchanged	Deteriorated
I	9	0	9
C	3	1	3
I-C	7	0	1

Note: The numbers for Conditions C and I-C are the numbers of groups. The numbers for Condition I are the number of participants.

Table 4. Number of participants' comments (% of all comments made at the session)

	I-C Condition (N= 8 groups)		C Condition (N= 7 groups)		I Condition (N= 18 participants)	
	Session 1	Session 4	Session 1	Session 4	Session 1	Session 4
Coherency	4 (14.3%)**	1 (4.5%)	4 (8.70%)+	0 (0.0%)	0 (0.0%)	0 (0.0%)+
Construction of the essay	2 (7.1%)	1 (4.5%)	10 (21.74%)	4 (19.05%)	6 (15.0%)	8 (15.4%)
Own conclusion	7 (25.0%)	7 (31.8%)	5 (10.87%)+	2 (9.52%)	9 (22.5%)	12 (23.1%)
Reader friendliness	8 (28.6%)	5 (22.7%)	4 (8.70%)*	4 (19.05%)	6 (15.0%)	15 (28.8%)+
Linguistic usage	1 (3.6%)+	2 (9.1%)	8 (17.39%)	10 (47.62%)**	3 (7.5%)	9 (17.3%)
Contents of the essay	6 (21.4%)	6 (27.3%)	15 (32.61%)	1 (4.76%)*	16 (40.0%)*	8 (15.4%)+
Total number of comments	28	22	46	21	40	52
Mean number of comments	3.50	2.75	6.57	3.00	2.22	2.88
Mean number of categories of comments	2.50	2.63	3.29	1.86	1.89	2.56

+: $p < .10$, *: $p < .05$, **: $p < .01$

Omnibus results First, we analyzed the comments made through all the four sessions. The mean total number of comments made by one group or one participant was calculated. In Condition C, it was 16.29 (SD=2.50). This indicated that in Condition C, one group made about 16 comments in the four EA sessions. By contrast, the mean total number of comments in Condition I-C was 11.88 (SD=3.48) and it was 10.06 (SD=3.455) in Condition I. A between-subject factor analysis of variance (ANOVA) was conducted and the result was significant ($F(2,30)=9.02$). The following multiple comparisons revealed that the differences between Conditions C and I-C and between Conditions C and I were significant. The participants in Condition C made significantly more comments than did those in the other two conditions.

However, when we compared the variety of the comments, we found that the opposite was the case. We counted the total number of categories of comments the participants mentioned at least once throughout the four sessions. In Condition I-C, the number of categories was 4.75 (SD=0.89), whereas it was 3.57 (SD=0.79) in Condition C. In Condition I, it was 4.11 (SD=0.96). The result of ANOVA was marginally significant ($F(2,30)=3.15$, $p=.06$) and the difference between Conditions C and I-C was significant ($p < .05$). When comparing Conditions C and I-C, the data supported our prediction. However, it was not clear what caused the performance to differ in Condition I-C and not in Condition I.

Comments made in Session 1 and 4 To examine the difference between the conditions more precisely, we investigated the comments made in the first (Session 1) and last (Session 4) sessions (Table 4). First, we compared the mean numbers of comments made in the first and last sessions. We conducted a 2x3 ANOVA with session and

condition as the between subject factors. The results revealed that the interaction effect reached significance ($F(2,30)=24.70$). A significant effect of session was found only in Condition C, in which the participants reduced the number of comments significantly. Moreover, ANOVA on the mean number of categories of comments showed that the interaction effect was significant ($F(2,30)=12.36$). The significant effect of session was found only in Condition C, showing that the participants produced less varied viewpoints in the last session than in the first session. By contrast, the effect of session was marginally significant in Condition I ($p=.09$). The participants in Condition I generated more diversity in viewpoint in the last session than in the first. Following this, a 6x6 chi-square test was conducted. The results reached significance ($\chi^2(25)=58.60$). The results of the residual analysis (Table 4) revealed that, in the first session, participants showed some bias in each condition, generating more or fewer comments for particular categories than expected. In the last session however, significant bias was seen only in Condition C. In addition, only marginally significant results were obtained in Condition I ($p < .10$).

In summary, the analysis of writing performance showed that only the participants in Condition I-C improved. However, the analysis of the comments revealed that the Conditions I and I-C did not differ. There were no significant differences between the two conditions regarding the variety and number of categories. The participants in both conditions made more varied comments than did those in Condition C. In addition, the participants in Condition I showed improvement with respect to developing diversity in their viewpoints through EA. Although the results of Conditions C and I-C supported our hypothesis, this cannot explain the results for Condition I.

General Discussion

This study investigated whether students' writing improved through an intervention designed to facilitate their monitoring activity. It was found that the participants developed more variety in their viewpoints of monitoring, and improved their writing performance through EA only when they worked in collaboration after working individually. The participants in Condition C showed more bias in their viewpoints, which hindered their improvement in writing. The results suggest that collaboration helps students enhance their viewpoints, but it is also necessary to provide them with the opportunity to prepare and clarify their own ideas individually.

Conversely, it was not clear why Individual EA, with instructions to learn from others' ideas, did not improve participants' writing performance. The participants achieved varied viewpoints, but their performance was not improved. We speculated that there was a difference in how they learned about the viewpoints of others. In Condition I-C, the participants needed to be involved in the formulation of viewpoints, whereas the participants in Condition I simply listened to the products of the other participants. Involvement in the formulation process appears to be important for putting the learnt viewpoints into practical use in writing. A more detailed analysis of the participants' learning process would be necessary in order to clarify the factors that contribute to the effects of the EA. Moreover, we need to look at how the participants' writing processes changed. As mentioned above, how the participants put the learnt viewpoints into actual writing situations differed. Although the change in participants was analyzed through the change in their comments, it would also be important to investigate the process of the participants on-line.

Note that the EA takes relatively little time and can be administered without much instructor scaffolding. Although our study did not compare EA with other teaching methods, the results indicate that this intervention functioned to improve students' writing performance, through encouraging their metacognitive monitoring. However, this does not mean that the EA should be implemented alone. The EA constitutes one possible intervention for improving student writing. It is well known that the essay-writing process is strongly affected by the context of writing, and expert scaffolding is effective (Galbraith & Rijlaarsdam, 1999). EA with an authentic context and scaffolding might lead to greater improvements in student writing. In addition, strategy teaching should be combined with the EA. Our study showed that the EA would work without these other techniques, but we do not deny the possibility of it being a better intervention when used with other teaching methods.

References

- Bereiter, C., Scardamalia, M., & Steinbach, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science*, 8, 173-190
- Breetvelt, I., Van Den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: When and how? *Cognition and Instruction*, 12, 103-123.
- Flower, L.S., & Hayes, J.R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg and E. R. Steinberg (Eds.), *Cognitive Processes in Writing*, (pp 31-50). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Galbraith, D., & Rijlaarsdam, G. (1999). Effective strategies for the teaching and learning of writing. *Learning and Instruction*, 9, 93-108.
- Hayes, J.R., & Flower, L.S. (1980). Identifying the organization of writing processes. In L. W. Gregg and E. R. Steinberg (Eds.), *Cognitive Processes in Writing*, (pp 3-30). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Janis, I. (1972). *Victims of groupthink*. Boston: Houghton Mifflin.
- Kellogg, R., T. (1988). Attentional overload: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 355-365.
- Kiyokawa, S., Okada, T., & Ueda, K. (2002). The effect of metacognitive suggestions on viewpoint change in collaborative problem solving. *Proceedings of the 4th International Conference on Cognitive Science*. [On CD-ROM].
- Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science*, 10, 151-177.
- Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 26, pp 125-141). New York: Academic Press.
- Okada, T. & Simon, H.A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science*, 21, 109-146.
- Sawyer, R.J., Graham, S., & Harris, K.R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self regulation: Effects on the composition skills and self efficacy of students with learning disabilities. *Journal of Educational Psychology*, 84, 340-352.
- Schwarz, B. B., Neuman, Y., & Bieauner, S. (2000). Two wrongs may make a right... If they argue together! *Cognition and Instruction*, 18, 461-494.
- Shirouzu, H., Miyake, N., & Masukawa, H. (2002). Cognitive active externalization for situated reflection. *Cognitive Science*, 26, 469-501.
- Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of text. *Journal of Educational Psychology*, 65, 66-73.
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1024-1037.
- Turner, M. E., & Pratkanis, A. R. (1998). Twenty-five years of groupthink theory and research: Lessons from the evaluation of a theory. *Organizational Behavior and Human Decision Processes*, 73, 105-115.
- Zimmerman, B.J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from Process to Outcome self-regulatory goals. *Journal of Educational Psychology*, 91, 241-250.

