

# The Effect of EFL Students' Self-Monitoring on Class Achievement Test Scores

Terry Vanderveen  
*Kanagawa University*

The effect of self-monitoring on an achievement test in an EFL classroom setting was investigated. The results of an experimental study that measured the difference in class achievement test scores between a self-monitoring and an external-monitoring group are reported. Participants were 114 first-year students in a Japanese university and college. Target language performance and behaviors were monitored and responses recorded on a protocol by the students in the self-monitoring group and monitored by the instructor in the external-monitoring group. The time required for implementing the protocol or treatment was also measured and was considered to be an important pedagogical factor. At an alpha level of .05, the difference in achievement scores between the groups was not statistically significant.

本研究は、外国語としての英語学習における自己モニターの効果について、到達度テストの結果を分析することにより検証した。日本の大学で英語を学習する1年生合計114名を学習者自身の自己モニターグループと教師による外部モニターグループの2つのグループに分け、到達度テストを実施し、差を検定した。それぞれ、言語使用行動を、前者のグループは自ら、後者のグループは教員が観察記録し、それをプロトコールとして分析した。その結果、 $\alpha$  0.05において2グループ間の間に統計的有意差は認められなかった。

**S**elf-regulated learning and learning strategy theory emphasize the role of self-observation and self-evaluation in learning. Arguments for the effectiveness of self-regulation are based on the assumption that conscious reflection on one's performance increases the frequency

and accuracy of target behaviors or performance. Self-regulation can be broken down into many subprocesses such as self-monitoring, self-instruction, self-evaluation, self-correction, self-reinforcement (Mace, Belfiore, & Hutchison, 2001), self-observation, self-judgment, and self-reaction (Zimmerman, 1989). The line separating these processes is often unclear (Benson, 2001) because as Schunk (2001) observes they are inter-related and dependent. This paper will use the term self-monitoring (SM) because it is the most widely used and, in my opinion, best describes the process discussed in this study.

O'Malley and Chamot (1990) define SM as "checking one's comprehension during listening or reading, or checking the accuracy and/or appropriateness of one's oral or written production while it is taking place" and contrast this with self-evaluation, which is "checking the outcomes of one's own language learning against a standard after the learning has been completed" (p. 232). Wenden (1991) offers similar definitions that differentiate SM and self-evaluation based on the time lapsed between production and assessment. This time period, however, does not change the nature of the evaluation process. Evaluation occurs both during and after learning. Therefore, for the purposes of the current study, SM and self-evaluation are considered a unitary construct referred to as SM and defined as conscious observation and evaluation, which is usually recorded, of one's own performance or behavior on a learning task.

The proposition that SM enhances learning has been widely discussed in general education as well as language education. SM has been investigated in relation to many aspects of learning, with behavior modification receiving the most attention. Researchers into behavior disorders in children have extensively reported improved classroom behavior or performance due to SM (Lam, Cole, Shapiro, & Bambara, 1994; Reid & Harris, 1993). However, not all studies of SM and behavior modification have yielded positive results. SM was ineffective in increasing productivity in a study of adults with mental retardation (Shapiro & Ackerman, 1983).

Mixed results are also reported from studies examining the impact of SM on academic outcomes. Mace and Kratochwill (1985) showed that SM significantly reduced L1 verbal nonfluencies in college students' speech, and Lan's (1993) experiment on the effects of SM on college students' statistics course grades resulted in the SM group outperforming the instructor-monitoring and control groups. However, in a similar experiment involving children studying math (Schunk, 1983), the posttreatment achievement scores of the SM group of children were comparable with those of the externally monitored group. SM also failed to improve per-

formance in a study of undergraduates' monitoring of success and failure in solving anagrams (Susser, 1981).

Whereas SM studies in education have been both behavioral and cognitive in nature, SLA researchers have focused solely on cognitive aspects. A cross-sectional speech monitoring study by van Hest (2000) shows that the ability to monitor one's L2 speech errors is a valid predictor of acquisition. In a review of monitoring and self-repair in L2 speech, Kormos (1999) concludes that the SM of L2 speech aids acquisition and is a positive correlate of proficiency. In two descriptive studies, Charles (1990) and Cresswell (2000) emphasize the importance of SM through the technique of writing notes or annotations. Checking one's reading comprehension is also considered useful and is recommended for developing reading skills (Block, 1992; Casanave, 1988). It has been extensively argued that SM and similar strategies are characteristics of good learners and enhance learning (Blanche & Merino, 1989; O'Malley & Chamot, 1990; Wenden, 1991). Despite these arguments for the effectiveness of SM, there has been little experimental evidence supporting claims that SM improves second language performance.

My search for studies that have manipulated SM and strategies similar to SM in the classroom yielded only two. Other studies offered anecdotal, theoretical, or correlational evidence. One of the empirical studies was by O'Malley, Chamot, Stewner-Manzanares, Russo, and Kupper (1985), who found statistically significant differences on speaking but not on listening posttests between treatment and control groups. The treatment group but not the control group was taught and directed to self- and peer-monitor their notes on the main points and cohesive markers of short speeches. However, only the treatment group was taught how to identify main points and cohesion markers. In other words, only the treatment group was explicitly taught structures and linguistic cues that would end up on the posttreatment test. This confounded strategy and structure instruction, strongly biasing the results in favor of a treatment effect. The time spent on strategy instruction and practice was 7 hours of class time during the fall semester. The total number of class hours in the semester was not mentioned in the study.

The treatment group in another experimental study (Viswat & Jackson, 1994) was also given more instruction than the control group, but in the form of "additional prelistening questions, which were designed to lead students to make use of the title, pictures, and information in the introductory statement to predict" (p. 241). Furthermore, significant teacher effects owing to different teachers teaching the strategies were

observed. Viswat and Jackson thought that this may have contributed to the large effect size of .57 in one of the treatment groups and felt that the study's "results [were] not conclusive" but "promising" (p. 247).

Oxford (1992) and O'Malley and Chamot (1990) have documented much of the theory and research into L2 learning strategies such as SM, but most of these rest solely on theoretical arguments. The lack of experimental investigations into SM, and the bias due to the treatment groups' added exposure to the target concepts, as in the two studies discussed above, points to the need for more research before claims can be made for the effectiveness of SM.

Lan (1993) and Schunk (1983) avoided the method bias present in the Viswat and Jackson (1994) and O'Malley et al. (1985) studies by providing external- or teacher-monitoring to the control group and comparing the results with those of the self-monitoring group. The control group was externally monitored by the instructor while each student in the treatment group monitored himself or herself, thus eliminating confounding effects due to one group receiving additional instruction. The present study, unlike previous monitoring or strategy instruction research in SLA, similarly controlled for bias arising from additional instruction.

Based on current SM theory, I assumed that the SM group would show greater improvement in class performance and thus score significantly higher on an end-of-term achievement test. To test this assumption, both groups were taught the same materials at the same rate and given the same test. The test scores were then compared to determine the impact of SM.

The time needed for the SM or strategy intervention was another variable considered. Time is an important factor in determining the efficacy of SM, yet in the SLA literature examined, the proportion of time required for SM or strategy training and administration was never reported. In a critical review of learner training, Rees-Millar (1993) addressed this issue, suggesting that if an excessive amount of time is required for strategy training, it may be better to use instructional resources for traditional tasks. The intervention in the O'Malley et al. (1985) study mentioned above required 7 hours in a semester. Although the total number of class hours in a semester was not mentioned, 7 hours is still a considerable amount of class time.

Time is an important factor in teaching decisions and needs to be included in discussions on the practical implications of SM intervention. Therefore, I measured the time required for the SM training and

implementation. This would not only provide an estimate of the time needed for SM, but also information on its relationship to the effect size of the intervention. To summarize, the following research questions were investigated in a tertiary education setting in Japan: (a) What is the effect of the SM of language-related behavior and performance on class test scores? (b) What is the amount of class time required to implement and employ SM?

## Method

### *Participants*

Each participant was a student at one of two schools: a Japanese co-ed university and a liberal arts college for women. The treatment group was comprised of two first-year English classes from the co-ed university and one first-year English class from the women's college. The control group was also drawn from two first-year English classes at the co-ed university and one first-year class from the college. All the courses were required, but the students had some choice of which section to enroll in. The women's college assigned students to classes based on their scores on a large standardized test, the General Tests of English Language Proficiency (G-TELP) (G-TELP Testing Services Center, 2002). Scores on this test for the treatment group were one standard deviation higher than the control group. As we shall see later, this difference had no bearing on the achievement test. All the students were in the humanities and relatively similar in academic achievement and educational background. None had had experience self-monitoring. For these reasons, the groups were considered equal for the purposes of the study.

There were 25, 21, and 17 students in the classes that made up the treatment group ( $n = 63$ ) and 26, 19, and 6 students in the control group ( $n = 51$ ) for a total of 114, excluding 6 students who did not write the test. From the women's college, there were 21 students in the treatment group and 19 in the control group.

### *Design and Procedure*

The SM protocol was designed for students in the treatment group to monitor their task behaviors and academic accuracy during class. Each student in the treatment group was given the protocol in the first class of the term. Course content and tasks were taken from *Fifty-Fifty Book One* (Wilson & Barnard, 1998), a beginning level grammar, speaking, and

listening text. The protocol items required the students to reflect, monitor, and record their task-related behaviors and outcomes. Behavior-related items that were monitored included (a) note taking of new vocabulary and structures, (b) the percentage of English spoken during speaking exercises, (c) the amount of eye contact during speaking exercises, (d) the students' effort, (e) the homework starting and finishing times, and (f) attendance.

Academic-related factors that were monitored included (a) the perceived difficulty of homework, (b) the homework score, (c) the students' perceived mastery of the homework, and (d) their perceived mastery of the text's grammar, listening, and speaking exercises (these three aspects were evaluated separately). The students also wrote the chapter number and the chapter's target structures (e.g. prepositions) on the protocol. Each student quantified his or her language-related behavior and academic performance as a percentage or as a point on a scale.

At the start of the course, the students were informed that the SM protocol would be evaluated solely on its completion and accuracy and would be worth 10% of the final grade. Points were not deducted from the protocol score if the homework was unfinished, or if the students failed to understand or achieve the target behaviors or tasks. However, they would lose points for falsely recording completion of their homework. I randomly checked approximately 70% of the protocols after every class and found that they were satisfactorily completed.

The same behavior- and academic-related factors that were self-monitored in the experimental group were also monitored by the control group. However, the control group's behaviors and outcomes were monitored by the teacher (i.e. externally). To externally monitor, the teacher randomly selected students and asked them questions regarding the same behavioral and academic items that were on the SM group's protocol. This was done at the beginning or end of class, depending on the specific item. For example, homework completion was monitored at the beginning of the class while perceived mastery of language structures covered in class was monitored at the end of the class. Due to time constraints, it was not possible to externally monitor the entire control group during every class, but all students were familiar with the monitoring items listed above and every student's work was monitored periodically. If they were not being monitored, they were observing the other students being monitored.

Both groups met once a week for 90 minutes and covered the same material during the same week over a 14-week term. Throughout the course the monitoring was consistent and proximal to the tasks, both

important considerations in SM theory (Schunk, 2001). Test achievement, the dependent variable, was measured one week after the term finished. Written test items were constructed to measure the students' knowledge of the speaking task target behaviors, vocabulary, and grammar structures covered during the course. There were 85 mostly closed items (three were open items related to direction and prepositions; there were no true/false or multiple choice items). The students took the same test at the same time at each institution. The control and experimental groups' test scores were compared to see how the students' self-monitoring and reflecting on the concepts studied and practiced in class affected their test scores.

### Results

After collecting and examining the protocols at the end of the course, I found that 95% of them had been properly completed. Average attendance for all the classes ranged from 84% to 92%. An achievement test was administered to both groups and its reliability was checked. The Kuder-Richardson 21 reliability estimate for the achievement test was .88, indicating consistent item performance. The control group's scores on the test were lower ( $M = 44.8$ ,  $SD = 12.49$ ,  $n = 51$ ) than the SM or treatment group's ( $M = 49.0$ ,  $SD = 12.13$ ,  $n = 63$ ). An independent samples  $t$  test assuming equal sample variances shows that at an alpha level of .05 the SM intervention did not result in a statistically significant difference between the two groups,  $t(112) = 1.80$ ,  $p = .07$  (two-tailed),  $d = .34$ . The 95% confidence interval of the difference of the means ranges from  $-8.76$  to  $.41$ . The dependent variables were normally distributed and Levene's test statistic ( $p = .89$ ) indicates homogeneity of variances.

As reported earlier, the G-TELP placement scores for the two classes in the women's college differed by an average of one standard deviation. The correlation between the G-TELP and the achievement test for the higher scoring class was  $r = .16$  and  $r = .20$  for the lower scoring class. These low correlations indicate that the two tests were unrelated and thus the one standard deviation difference in the G-TELP scores of the two women college classes at the beginning of the course should not have affected the class test scores. Any differences were likely due to differences between the test types, the G-TELP being a norm-referenced test measuring proficiency and the classroom test being a criterion-referenced test measuring topics covered over the length of the course. The SM group required 30 minutes training during the first class and approximately 12 to 15 minutes for completing and checking the protocol in each of the

subsequent classes. Over a 14-class, 21-hour term, this required approximately 4 hours or 20% of class time.

### Discussion

The main purpose of this study was to determine the impact of SM on class achievement scores. The data in this and other empirical studies in both general education (e.g. Schunk, 1983; Shapiro & Ackerman, 1983; Susser, 1981) and SLA (e.g. Viswat & Jackson, 1994; O'Malley et al., 1985) support the hypothesis that SM has no significant effect on achievement. However, the confidence intervals and effect size ( $d = .34$ ) show that there is a .95 probability that a small positive effect in addition to no effect may also be observed in the population.

The present study also set out to measure the amount of time required for the SM intervention. Perhaps more important than the failure to reject the null hypothesis—that SM intervention is of no use—was the approximately 4 hours or 20% of class time required for implementing and using the protocol. This result is consistent with the 7 hours needed for training in the O'Malley et al. (1985) study. In practical terms, educators considering SM or other metacognitive strategy instruction should anticipate the possibility of large time costs against minor or no increases in student achievement.

The sample size was admittedly small. However, the study was based on current theoretical arguments that SM causes significant changes in achievement. If this had been true, the sample size should have been sufficient. Assuming SM intervention results in effect sizes the same or greater than the current study, replication studies would need a sample larger than 270 students ( $n_1, n_2 > 135$ ).

My concern that the higher standardized test scores (G-TELP, 2002) of part of the treatment group would influence their class test scores was alleviated by the extremely low correlations between the tests. In other words, the scores of the class achievement test were not influenced by differences in the institutional placement test scores of the students. The combination of behavioral and cognitive outcomes in the protocol was not a limitation, but measuring the separate effects of these two factors might tell us if one has a greater impact on achievement.

The sheer number of arguments for SM and strategy instruction give the appearance that they have been extensively researched, yet two important strategy researchers, O'Malley and Chamot (1990) acknowledge that "there has been little confirmation of the effectiveness of strategy



training with second language tasks" (p. 224) and to date there still has not been much empirical research. The results of the current experimental study also cannot confirm the effectiveness of SM strategy training.

The findings presented here indicate that further empirical investigations into the impact and time requirements of SM are needed before accepting the hypothesis that SM training is effective and practical in the classroom. Teachers must realize the possibility that there may be little or no effect from SM and that a considerable amount of time may be needed to implement it in the classroom.

### Acknowledgments

*I would like to thank the editor and the two anonymous referees of JALT Journal for their advice on the revision of this paper.*

Terry Vanderveen teaches at Kanagawa University. His research interests include motivation and evaluation. He can be contacted at <teratkana@yahoo.co.jp>.

### References

- Benson, P. (2001). *Teaching and researching autonomy in language learning*. Essex, England: Pearson Education.
- Blanche, P., & Merino, B. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39, 313-340.
- Block, E. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly*, 26, 319-343.
- Casanave, C.P. (1988). Comprehension monitoring in ESL reading: A neglected essential. *TESOL Quarterly*, 22, 283-302.
- Charles, M. (1990). Responding to problems in written English using a student self-monitoring technique. *ELT Journal*, 44, 286-293.
- Cresswell, A. (2000). Self-monitoring in student writing: Developing learner responsibility. *ELT Journal*, 54, 235-244.
- G-TELP Testing Services Center. (2002). *General tests of English language proficiency*. La Jolla, CA: Author.
- Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning*, 49, 303-342.
- Lam, A., Cole, C., Shapiro, E. S., & Bambara, L. M. (1994). Relative effects of self-monitoring on-task behavior, academic accuracy, and disruptive behavior in students with behavior disorders. *School Psychology Review*, 23, 44-58.
- Lan, W. Y. (1993). The effects of self-monitoring on students' course performance, use of learning strategies, attitude, self-judgment ability, and knowledge representation. *The Journal of Experimental Education*, 64, 101-116.

- Mace, F. C., Belfiore, P. J., & Hutchison, J. M. (2001). Operant theory and research on self-regulation. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 39-65). Mahwah, NJ: Erlbaum.
- Mace, F. C., & Kratochwill, T.R. (1985). Theories of reactivity in self-monitoring: A comparison of cognitive-behavioral and operant models. *Behavior Modification*, 9, 323-343.
- O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Russo, R., & Kupper, L. (1985). Learning strategy applications with students of English as a second language. *TESOL Quarterly*, 19, 557-584.
- Oxford, R. L. (1992). Research on second language learning strategies. *Annual Review of Applied Linguistics*, 13, 175-187.
- Rees-Millar, J. (1993). A critical appraisal of learner training: Theoretical bases and teaching implications. *TESOL Quarterly*, 27, 679-689.
- Reid, R., & Harris, K. R. (1993). Self-monitoring of attention versus self-monitoring of performance: Effects on attention and academic performance. *Exceptional Children*, 60, 29-40.
- Schunk, D. H. (1983). Progress self-monitoring: Effects on children's self-efficacy and achievement. *Journal of Experimental Education*, 51, 89-93.
- Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In B.J. Zimmerman & D.H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 125-151). Mahwah, NJ: Erlbaum.
- Shapiro, E. S., & Ackerman, A. (1983). Increasing productivity rates in adult mentally retarded clients: The failure of self-monitoring. *Applied Research in Mental Retardation*, 4, 163-181.
- Susser, S. H. (1981). *Self-monitoring success and failure: Evidence for a mediating mechanism*. Paper presented at the annual convention of the American Psychological Association, Los Angeles, CA. (ERIC Document Reproduction Service No. ED217348)
- van Hest, E., (2000). Analyzing self-repair: An alternative way of language assessment. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 75-88). Mahwah, NJ: Erlbaum.
- Viswat, L.J. & Jackson, S.A. (1994). The effect of strategies training on student errors on a listening cloze. *Journal of the Himeji Dokkyo University College of Foreign Languages*, 235-249. (ERIC Document Reproduction Service No. ED411669)
- Wenden, A. (1991). *Learner strategies for learner autonomy*. London: Prentice Hall International.
- Wilson, W., & Barnard, R. (1998). *Fifty-fifty book one*. Singapore: Pearson Education.
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81, 329-339.