THE USE AND EVALUATION OF STUDY GUIDES WITH MIDDLE SCHOOL STUDENTS¹

MARLENE FARNUM AND THOMAS A. BRIGHAM

WASHINGTON STATE UNIVERSITY

Two experiments were conducted with 24 fifth-grade students participating. In the first experiment, quiz performance with study guides was compared to quiz performance without study guides. The group whose students received study guides always scored higher than the group who did not receive study guides. Experiment II assessed the feasibility of students scoring their own study guides. The reliability of student scoring was found to be 93% over all study guides and there were no significant differences between "self-scored" and "teacher-scored" groups on the weekly quizzes.

DESCRIPTORS: academic behavior, study guide effects, multiple baseline, teacher evaluation, self-evaluation, fifth-grade students

In an analysis of research related to Personalized System of Instruction (PSI) procedures, Lloyd (1977) concluded that study guides may be one of the critical independent variables in increased performance when PSI procedures are compared with traditional lecture procedures. For example, Semb, Hopkins, and Hursh (1973) found that performance increased on criterion tests as a function of items appearing on study guides and unit tests. They also compared correct responses on unit tests when students were given study guides with answers, study guides with no answers, and no study guides. Test scores were higher in the two study guide conditions. Next, Semb (1975) looked at the effect of unit tests on final examination scores. All students in this experiment completed study guides, but some took unit tests and others did not. No reliable score difference was found between students who did and those who did not take the unit quizzes, indicating that working on study guides is an educationally important activity in and of itself.

In other research with study guides, Whitehurst (1972) demonstrated that errors are reduced on unit tests when the answers to study-guide questions are written.

This paper describes two experiments that examined various aspects of using small unit study guides in a fifth-grade social studies class. Extrapolating from the work done with college students, the first study was designed to test the hypothesis that study guides would improve the test performance of middle school students. The second experiment examined the effects of having students act as their own proctors by scoring their own work.

EXPERIMENT I

METHOD

Subjects and Setting

Ten boys and 14 girls in a regular fifth-grade classroom, located in a middle-class neighborhood, participated in the study. Research in other curriculum areas was being conducted simultaneously with the present research. For example, in mathematics the effects of teacher-set weekly page goals and student-set weekly page goals were compared with regard to rate of problems worked, per cent correct, per cent of weekly goals reached, and differences between teacher-

¹This research was supported by Grant 1 RO1 MH26084-01 from the Applied Branch of NIMH to the Center for Applied Behavioral Research, Washington State University. Reprints may be obtained from Thomas A. Brigham, Department of Psychology, Washington State University, Pullman, Washington 99164.

set and student-set goals. The students were also operating under a reinforcement program for work completed in reading.

The class was divided into two experimental groups on the basis of seating arrangement in the classroom. Normally, six students comprised one work area. The work areas were separated by partitions, so that the students in one work area were not exposed to students in the other work areas. Each experimental group occupied two separate work areas. This division reduced the possibility of the student being exposed to the study guides of the other group.

Curriculum

The curriculum consisted of two standard social studies texts. Text A, *Asking about the U.S.A. and Its Neighbors* (Davis and Hunkins, 1971), was from the *Exploring the Social Sciences Series* (1971), and contained information on the U.S.A. and other countries. Text B, *I Earn, Explore and Excel* (Glynn, Thomas, and Shee, 1973), was part of the *Triple "I" Series* (1970), and its major purpose was to expose the students to different values, interpersonal relationships, and to "aid in the development of self-knowledge". The texts were used concurrently in the classroom. Both texts were already divided into small units of approximately equal page length. The study guides and weekly quizzes were constructed by the classroom teacher

and a research associate with a background in test construction.

Dependent Variables

The major dependent variable was a difference score computed between the two groups on weekly quizzes over the same text material in the same textbook. The use of a mean difference score controlled for changes in the difficulty of the material across units. For example, in those cases where the material covered or the weekly quiz was less difficult, quiz scores would increase for both groups. However, the differences between the group scores would still be present in those cases in which the experimental manipulation was effective.

Procedure

A complex multiple-baseline semi-reversal design, across groups and social studies texts, was employed (see Table 1). This design made it possible to control for variations in difficulty of the curriculum materials, and to make both within-group and between-group comparisons.

Baseline

In the three-week baseline phase, all students were assigned the same units to read in the two social studies books. On Monday of each week, quizzes over the assigned units in both books were administered to all students. The quizzes

Table 1
Sequence of Group 1 and Group 2 Using Study Guides, Experiment I

	<i>Baseline Condition 1— Three Weeks</i>	<i>Study Guides Condition 2— Three Weeks</i>	<i>Reversal Condition 3— Three Weeks</i>
BOOK A			
Group 1	No Study Guides	No Study Guides	Study Guides
Group 2	No Study Guides	Study Guides	No Study Guides
BOOK B			
Group 1	No Study Guides	Study Guides	No Study Guides
Group 2	No Study Guides	Study Guides	Study Guides

consisted of short-answer essay questions. The maximum number of possible points on the quizzes for Text A ranged from 10 to 14, with a mode of 10; the range for Text B was 8 to 12 points, with a mode of 10. Two points were awarded for each correct answer, and one was given if an answer was partially correct. There was no minimum criterion on quizzes, nor were there any make-up quizzes. Quizzes were graded by the experimenter after the social studies class ended and returned to the students during the following social studies period. The points earned by the students were used to determine their class grade, but there were no other systematic consequences for high or low scores. An assistant unfamiliar with the research rescored all quizzes. Reliability was calculated by the following formula:

$$\% \text{ agreement} = \frac{\# \text{ agreements on items}}{\# \text{ agreements of items} + \# \text{ disagreements on items}} \times 100.$$

Study Guides I

For the next three weeks, Group 1 received study guides for Text B, but did not receive study guides for Text A. In this same phase, Group 2 received study guides for Text A, but not for Text B. The same units were read by both groups. This procedure allowed a comparison of groups on the Monday quizzes within each social studies book, thus controlling for any changes in the difficulty of the unit material. Study guides consisted of fill-in-the-blank statements that covered the unit material. Each blank counted as one point. The number of blanks per study guide ranged from 14 to 20, with a mode of 16 for Text A. For Text B, the range was from 10 to 20, with a mode of 14. Each study guide was scored for each student and the number of points was recorded at the top of the study guide. These points were not used for grades; rather, they indicated to the students only how they had performed on the study guide. If a student did not finish the study guide or was absent, his or her name was put on the black-

board with a due date. This was standard classroom procedure for all incomplete work for all academic areas. The teacher checked at the end of each day to see that all assignments had been completed. If a student was absent or a study guide was not completed, the student was required to stay in at recess or after school and finish the assignment. There were no other consequences. The study guides were graded by the first author after the social studies class and returned to the students the following day. All study guides were rescored by an undergraduate research assistant.

The quizzes followed the same format as in baseline. The quizzes were graded by the first author after the social studies class. Reliability was conducted in the same manner as in baseline.

Reversal of Study Guides

For the next three weeks, Group 1 did not receive study guides for Text B but did use study guides for Text A, while Group 2 switched from study guides for Text A to study guides for Text B. A difference score was again computed between groups within each social studies text.

RESULTS AND DISCUSSION

Student performance with the study guides ranged from 88% to 95% for Text A, and from 87% to 95% for Text B. There were nine instances in Experiment I in which students were absent for an entire week, thus not completing the study guides and quizzes for that week. Reliability for the experimenter scoring of study guides was 99% for Text A and 98% for Text B; reliability for weekly quizzes was 92% for Text A and 97% for Text B.

The test scores during the baseline phase (see Table 2) indicate that Text A was more difficult for the students. During baseline, scores for Text A ranged from 20% to 42%; scores for Text B were from 50% to 75%. These substantial differences in per cent correct on unit quizzes for the two books persisted throughout the study.

Fig. 1. Mean difference in per cent correct: Group I minus Group II.

The dependent variable was the mean difference in per cent correct between Group 1 and Group 2 on the Monday quizzes. Figure 1 shows that Group 1 scored consistently higher on the quizzes given during baseline in Book A. The differences favoring Group 1 of 12%, 22%, and 4% on the first three quizzes indicate that the assignment to groups by seating arrangement did not evenly match the students on this book.

When Group 2 received the study guides in the next condition, however, they scored higher on all quizzes, even though they were less skilled than Group 1 in this book during baseline. The differences favoring Group 2 were 21%, 13%, and 25% on these quizzes. A *t*-test between Groups 1 and 2, on the average number of correct answers per condition, yielded a *t* of 2.73 (*df* = 22, *p* < 0.01). When in the final condi-

tion Group 1 received the study guides for Book A, Group 1 was again superior to Group 2 on three quizzes by 29%, 29%, and 23% ($t = 6.36$, $df = 22$, $p < 0.001$). Note that the magnitude of the differences between Group 1 and 2 is higher in the study-guide condition than for the baseline condition.

Similar differences were found with Book B. Figure 1 shows that the initial mean differences between groups with Book B is small, with Group 1 scoring slightly higher on all baseline quizzes. When study guides were instituted for Group 1, the differences scores for this group rose above baseline for three consecutive quizzes to levels of 10%, 10%, and 30% ($t = 3.14$, $df = 22$, $p < 0.01$). When the condition was reversed, Group 2 consistently scored higher than Group 1 by difference scores of 14%, 22%, and 6% ($t = 3.18$, $df = 22$, $p < 0.01$).

Overall, the group that used study guides for the given text scored higher on weekly quizzes (see Table 2). However, given the variability in the difficulty of the unit material, not all weekly quizzes in the study-guide conditions were greater than baseline scores. This points to the

importance of using the difference score on the weekly quizzes as the dependent measure.

The results from Experiment I demonstrate that study guides can be used to improve the work of middle school children as well as college students. The use of study guides always improved the performance of students on the weekly quizzes. This was true for both the within-group comparisons, which controlled for students' abilities, and the between-group comparisons, which controlled for difficulty of the course material. However, for a procedure to be practical as well as effective it must be possible to implement it in the standard self-contained classroom. Since McLaughlin and Malaby (1975) demonstrated that sixth-grade students could successfully function as proctors, it was decided to investigate a similar student evaluation program. Our preliminary studies (Farnum, Note 1; Niemann and Brigham, Note 2) have shown that elementary school students could accurately evaluate their own work. Based on this pilot research, a procedure that allowed students to score their own study guides was developed and evaluated in Experiment II.

Table 2
Group Per Cent Correct on Weekly Quizzes, Experiment I

Book A, Group 1								
Baseline			No Study Guide			Study Guide		
42%	42%	36%	44%	42%	21%	44%	52%	54%
Book A, Group 2								
Baseline			Study Guide			No Study Guide		
30%	20%	32%	65%	55%	46%	16%	22%	31%
Book B, Group 2								
Baseline			Study Guide			No Study Guide		
75%	75%	55%	60%	90%	76%	75%	50%	50%
Book B, Group 1								
Baseline			No Study Guide			Study Guide		
72%	72%	50%	50%	80%	46%	89%	72%	65%

EXPERIMENT II

METHOD

Subjects and Setting

The subjects, setting, and curriculum material were the same as in Experiment I. Experiment II was conducted immediately after Experiment I was completed.

Procedure

A multiple-baseline semi-reversal design was employed to assess differences on weekly quizzes as a function of student or teacher correction of the weekly study guides (see Table 3).

For the first three sessions, each student received a study guide with his/her name, and "teacher correct" or "student correct" written in the upper right-hand corner. For Text A, the "teacher correct" group consisted of the students in Group 1. The students in Group 2 corrected their own study guides. After finishing their study guides, students in the "teacher correct" group took their papers to one of two research assistants. The study guides were corrected with the student present and the correct answer was written in by the research assistant for any missed items. Students in the "student correct" group took their study guides to a special table, where

answer sheets and special marking pens were located. They marked their answers correct or incorrect and were instructed to fill in any answers that had been omitted or that were incorrect. The corrected study guides were given to the research assistants. There were no contingencies for student grading of the study guides. The number of students who failed to write in the correct answer or failed to correct the study guide was tallied. Study guides were completed on Wednesday and Thursday, and on the following Monday a quiz was given covering the unit material. For Text B, the "student scoring-teacher scoring groups" were reversed. The mean group per cent correct was compared.

*Reversal of Teacher Correct
Compared with Student Correct*

For the next three sessions, Group 1 corrected their own study guides in Text A; for Group 2, study guides were corrected by the research assistants. Again, the opposite procedure was used with Text B.

Preference

For the final three weeks of the study, when the students received their study guides on Wednesdays, they were individually asked if they

Table 3
Sequence of Group 1 and Group 2 for student correction and teacher correction of study guides Experiment II.

		<i>Teacher Correct versus Student Correct Condition 4— Three Weeks</i>	<i>Reversal Condition 5— Three Weeks</i>	<i>Preference Condition 6— Three Weeks</i>
BOOK A	Group 1	Study guides Teacher correct	Study Guides Student correct	Study Guides Preference correcting
	Group 2	Study guides Student correct	Study guides Teacher correct	Study guides Preference
BOOK B	Group 1	Study guides Student correct	Study guides Teacher correct	Study guides Preference
	Group 2	Study guides Teacher correct	Study guides Student correct	Study guides Preference

would prefer to correct their own study guides or have a research assistant correct their study guides. The student's preference was noted at the top of the study guides for both books. The procedures for correcting the study guides were the same as in the preceding two conditions.

Reliability

Reliability was assessed for the student-corrected and teacher-corrected study guides, and for the experimenter-scored quizzes, by several assistants not familiar with the research. All agreements and disagreements were tallied and reliability was computed as in Experiment I.

RESULTS AND DISCUSSION

When weekly scores on quizzes were compared as a function of correcting method, no systematic differences were found (see Table 4). That is, the scores on the weekly quizzes were approximately equal, regardless of who corrected the study guides. Student scoring of study guides was reliable. For the 260 study guides completed, only 17 items were scored incorrectly by the students. The per cent correct on study guides ranged from 83% to 95% for Text A, and 75% to 95% for Text B. Reliability for the experimenter scored quizzes was 93% for Text A and 95% for Text B.

The last condition of this experiment assessed the students' preference for scoring their own study guides. In the first week, 20 of the 24 students elected to score their own study guides, 22 of 24 elected to score their study guides the second week, and 20 of 24 the last week. Because of

the students' overwhelming preference for student-correcting procedures, the mean per cent correct scores for the preference condition are not meaningful as comparison scores. However, again there did not appear to be any real difference between the test performance of the large "student-correct group" and the few "teacher-scored" students. Similarly, overall test performance remained high in Text B and reasonably good in Text A. Finally, the students' preference for and accuracy at self-scoring suggests that it is a procedure that teachers could easily implement in self-contained classrooms.

SUMMARY

Study guides were shown to be educationally effective in Experiment I. There was never a case in which the group whose students received study guides scored lower than the group without study guides on the weekly quizzes. Next, Experiment II demonstrated that students were willing to score their own study guides, and that they did so accurately. As a consequence, the overall study guide-self-scoring procedure is educationally effective and practical as well.

The success of this limited project suggests that one of the potential routes for improving instruction in elementary and middle school classrooms is to combine PSI procedures with self-management programs. PSI techniques have been demonstrated to be powerful educational procedures, but they need much manpower to run them (Pennypacker, Heckler, and Pennypacker, 1977). On the other hand, research in

Table 4
Group Per Cent Correct on Weekly Quizzes, Experiment II

	<i>Tests</i>					
	1	2	3	4	5	6
BOOK A						
Teacher Correct	54%	68%	52%	43%	63%	49%
Student Correct	44%	66%	54%	54%	65%	53%
BOOK B						
Teacher Correct	88%	70%	82%	81%	91%	83%
Student Correct	85%	79%	71%	78%	86%	88%

classroom management using self-control or self-management programs (e.g., Glynn, Thomas, and Shee, 1973) indicates that students can be taught to make a major contribution to the operation of their own classroom. Bushell and Bushell (1976) described a classroom system where students managed their own time schedules. Similarly, in the present study students were taught to self-score their own quizzes to handle some of the time-resource problems. It is likely that programs can be developed so that students can be systematically taught to manage their own PSI courses.

REFERENCE NOTES

1. Farnum, M. J. *The effects of self-evaluation procedures on academic performance and attitudes of middle school students*. Unpublished Masters Thesis, Washington State University, 1975.
2. Niemann, J. and Brigham, T. A. *Mathematical achievement under conditions of choice: teacher-evaluation vs. student-evaluation*. Paper presented at the Washington Education Research Association Conference, May, 1975.

REFERENCES

- Bushell, S. A. and Bushell, D., Jr. A dual contingency procedure to support student self-determination and performance. In T. A. Brigham, R. Hawkins, J. Scott, and T. McLaughlin (Eds), *Behavior analysis in education: Self-control and reading*. Dubuque, Iowa: Kendall/Hunt, 1976. Pp. 3-14.
- Davis, O. L., Jr. and Hunkins, F. D. *Asking about the U.S.A. and its neighbors*. New York: American Book Company, 1971.
- Franco, J. M., Rydell, W., and Padelford, W. B. *I earn, explore and excel*. New York: American Book Company, 1970.
- Glynn, E. L., Thomas, J. D., and Shee, S. M. Behavioral self-control of on-task behavior in an elementary classroom. *Journal of Applied Behavior Analysis*, 1973, 6, 105-133.
- Keller, F. S. New reinforcement contingencies in the classroom? *American Psychologist*, 1965, 20, 542.
- Lloyd, K. E. Behavior analysis and technology in higher education. In T. A. Brigham and A. C. Catania (Eds), *Handbook of applied behavioral research*. New York: Irvington Press/John Wiley & Sons, 1977. Pp. 313-358.
- McLaughlin, T. F. and Malaby, J. E. Elementary school children as behavioral engineers. In E. Ramp and E. Semb (Eds), *Behavior analysis: areas of research and application*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. 1975. Pp. 319-328.
- Pennypacker, H. S., Heckler, J. B., and Pennypacker, S. F. The Personalized Learning Center: A university-wide system of personalized instruction. In T. A. Brigham and A. C. Catania (Eds), *Handbook of applied behavioral research*. New York: Irvington Press/John Wiley & Sons, 1977. Pp. 591-617.
- Semb, G., Hopkins, B. L., and Hursh, D. E. The effects of study questions and grades on student test performance in a college course. *Journal of Applied Behavior Analysis*, 1973, 6, 631-642.
- Semb, G. An analysis of the effects of hour exams and student-answered study questions on test performance in a course taught by personalized instruction. In J. M. Johnson (Ed), *Behavior research and technology in higher education*. Springfield, Illinois: Charles C Thomas, 1975.
- Whitehurst, G. J. Academic responses and attitudes engendered by a programmed course in child development. *Journal of Applied Behavior Analysis*, 1972, 5, 283-291.

Received 23 August 1976.

(Final Acceptance 2 September 1977.)