*EFFECTS OF TRAINING STUDENTS TO IDENTIFY
THE SEMANTIC BASE OF PROSE MATERIALS*

JOHN A. GLOVER, JOHN W. ZIMMER, ROBERT W. FILBECK,
AND BARBARA S. PLAKE

THE UNIVERSITY OF NEBRASKA AND THE UNIVERSITY OF TOLEDO

Feedback and feedback plus points toward a course grade were applied to the attentional behaviors (defined as the ability to identify the semantic base of text passages) of 30 undergraduate students participating in a reading comprehension development program. Correct underlining was increased, extraneous underlining was decreased, and post-reading comprehension test scores improved as a result of the procedures. Scores on a standardized test of reading comprehension also increased significantly.

DESCRIPTORS: reading, recall, information processing, feedback, college students

With the continuing and intensifying emphasis on prose materials in our educational system (e.g., "back to basics"), educators could derive considerable benefit from information on procedures to enhance student comprehension of written materials. At present, a body of research, based in part on Ausubel's (1960) work on advanced organizers (i.e., highly general organizing materials presented prior to a reading passage designed to facilitate learning of the material) and Rothkopf's (1966) analysis of mathemagenic activities (activities which give "birth to learning"), has been devoted principally to devising aids (objectives, questions inserted in text, and advanced organizers) for increasing student comprehension of reading materials. However, the promise of adjunct aid

research as yet remains unfulfilled, although early reports of the use of adjunct aids were almost uniformly positive. Faw and Waller's (1976) extensive critique of the literature, for instance, expressed considerable concern that only two studies with objectives (Gagne & Rothkopf, 1975; Jenkins & Deno, 1971), two with inserted questions (Hiller, 1974; Hunkins, 1969) and one with advance organizers (Clawson & Barnes, 1973) used true control groups and held reading time equal for both experimental and control groups. Additionally, these studies did not obtain significant effects on overall comprehension for groups using these adjunct aids. These countervailing reports provided an impetus for further analyses of the role and function, if any, of adjunct comprehension aids.

Recently, a series of studies (Glover, Zimmer, & Bruning, 1980; Zimmer, Petersen, Ronning, & Glover, 1978; Glover & Zimmer, Note 1; Zimmer, Note 3; Zimmer, Note 4) demonstrated that adjunct aids have no significant effect on reading comprehension regardless of the "level of learning" (i.e., knowledge, comprehension, application) required by the aids or the nature of the dependent measure used. The results of these studies led to subsequent, "finer grain" analyses of the failure of adjunct aids to affect comprehension test scores, providing some tenta-

The authors are indebted to Theresa Glover for her work in preparing materials for the study. Additionally, several persons helped with the development of materials and scoring. Thanks are extended to Roger H. Bruning, Royce R. Ronning, Barry Roberts, Chip Borgstadt, and John Ingram. We are also grateful to Kris DeVries and Deb Olson for their efforts on our behalf. Reprints may be requested from John W. Zimmer, Department of Educational Psychology, University of Toledo, 2801 W. Bancroft, Toledo, Ohio 43606 or from any of the other authors at the Department of Educational Psychology and Measurements, University of Nebraska, Lincoln, Nebraska 68588.

tive reasons why they might not have worked: (a) students do not obtain feedback for the use of adjunct aids during reading, (b) there is no reinforcement for the effective use of aids, (c) knowledge of results on reading comprehension are long delayed, (d) students may simply not know how to use adjunct aids, (e) adjunct aids are often superfluous to the major ideas contained in prose materials, and (f) adjunct aids are often not designed to help readers develop reading skills; instead, they appear to be selected for facilitating test construction or other unrelated considerations.

For these and other reasons, research on adjunct aids is unlikely to prove beneficial for either readers or practitioners. Procedures intended to modify prose materials and subsequently affect reading comprehension without consideration of the reader or the effects of the modifications on the reader grossly neglect the lessons of applied behavior analysis. However, some of the analytical and methodological techniques developed in adjunct aid research may prove highly appropriate to applied behavior analysts examining reading comprehension and allied areas.

An applied behavior analysis approach to reading comprehension must include an identification of reading comprehension behaviors. Reading comprehension is typically assessed via postreading tests (Petersen, Glover, & Ronning, 1980) McCrudden (Note 2), however, points out that such test results do not allow us to infer the existence of any *specific* reading behaviors. That is, these tests measure "final" comprehension, not behaviors which lead to comprehension. One method for analyzing relatively complex behaviors (with considerable currency in the applied behavior analysis literature), is task analysis (cf. Resnick & Beck, 1977). Briefly, task analytic (TA) procedures involve the delineation of component steps, level of learning, and functional prerequisites of certain target objectives (Gagne, 1977). For many instructional activities where fundamental interrelationships are known (at least theoretically, as in the case of

mathematics) TA processes are an exceedingly valuable adjunct to the teaching process. However, with our presently limited knowledge of the reading process, TA would seem to contribute primarily to the arrangement of instructional events and not to an explication of the more fundamental processes underlying reading. A more advantageous approach to the preliminary assessment of reading comprehension behavior would seem to be provided by Kintsch's (1978) suggestion that any text passage can be viewed at two levels, surface features and the semantic base. Although knowledge of surface feature components are important for general writing and communication skills, it is most often the semantic base of reading materials to which we want our readers to attend. That is, we want readers to be able to make verbal or written responses indicating that they have identified and can paraphrase the "major ideas" in a prose array.

Given that attention to the semantic base of text materials is an integral part of reading comprehension (e.g., Kintsch, 1978; Smith, 1978), procedures that identify and bring attentional behaviors under experimental control ought to enhance the reading comprehension of students. Previous research (Glover, Zimmer, & Bruning, 1980; Petersen *et al.*, 1980; Zimmer *et al.*, 1978) has indicated that student underlining of material in prose arrays can be an effective measure of student attentional behaviors. Zimmer *et al.* (1978), for example, determined that student underlining responses could be used to determine which portions of prose arrays were attended to by students. They also found that accurate underlining of the semantic base of the text was significantly related to correct responding on posttest comprehension items covering that specific material.

The current study reports an attempt to bring student attentional behaviors, defined as correctly underlining the semantic base of prose passages, under experimental control and to relate them to posttest and standardized measures of reading comprehension.

METHOD

Participants were 30 undergraduate education majors enrolled in an educational psychology course. Participants were told the program was an experimental technique designed to increase reading comprehension. There were 9 females and 21 males, ranging in age from 18 to 26 years. The mean scores of participants on the comprehension section of the Nelson-Denny (1973) reading test was the 18th percentile with scores ranging from the 10th to the 24th percentile. Grade point averages of participants ranged from 2.06 to 2.56 out of a possible 4.00 with a mean grade point average of 2.43.

All data were collected in individual sessions in a small conference room set aside for that purpose. Student sessions were determined by scheduling convenience and, hence, occurred at somewhat different times across days for a period of 7 weeks.

Materials

Alternate forms of the comprehension section of the Nelson-Denny reading test (1973) were used as pre- and posttests with the sequence of the alternate forms randomly determined for participants, i.e., some received form A first and others received form B first, with the alternate form then used as the posttest. This test was selected for its significant predictive value for reading comprehension (Zimmer, Glover, Ronning, & Petersen, 1979).

The reading materials used in this study consisted of a series of 30, 800-word essays previously used in reading research (e.g., Glover, Ronning, Petersen, & Zimmer, 1980; Petersen et al., 1980; Zimmer et al., 1978; McCrudden, Note 2, Zimmer, Note 3, Zimmer, Note 4). These materials were from three different areas, the natural history of canines, the physical development of felines, and descriptive astronomy. All essays were at the college sophomore level of difficulty (c.f., Zimmer et al., 1978). The essays were sequenced in random order for each participant, with the materials typed on rolls of

label paper. The label paper ($1\frac{1}{2}'' \times 6\frac{1}{2}''$; 3.8 cm $\times$ 16.5 cm) was plain and could be removed from its backing by peeling up the paper at one corner, revealing the backing. Approximately five double-spaced lines of roughly nine words each could be typed on a sheet. Paragraphs were thereby not adhered to in the preparation of the text material and there was a variance from 34 to 51 words in the 5-line units, with a mean of 41 words and 3.24 sentences per unit.

During baseline conditions participants were instructed not to peel the paper, and the backing did not contain any information. However, during the experimental phases the backing had the semantic base of the 5-line unit printed on it so that participants could see these by peeling off the label paper.

General Procedures

At the beginning of each session, participants were asked to carefully read an 800-word passage, and to underline the minimum number of words indicating the meaning (semantic base) of each 5-line unit. Students were given 5 min for reading and studying the passage, after which they completed a brief 10-item multiple-choice test over the content. Five minutes were allotted for the test taking. Tests were scored immediately, with results recorded and provided to participants.

The student underlining was scored overnight against standards developed by the raters. Participants received the essay read during the previous session when they returned for their next session with the comment "good" written on it. During experimental conditions nothing was written on the essays while posttest procedures remained the same.

Experimental Conditions

Baseline. During baseline periods, the general procedures described above were in effect.

Immediate feedback. During the immediate feedback for underlining condition, students were instructed to underline the least possible number of words in each 5-line unit which,

when taken together, would express the meaning (semantic base) of that unit. They were also instructed to peel the label off immediately after they had finished and compare the words they had underlined to those printed by the experimenters on the backing paper. Other procedures remained identical to baseline conditions.

Immediate feedback plus points. Part of each student's classroom assignment was a study skills unit worth 100 points. During the feedback plus points condition, students were again instructed to underline carefully the minimum number of words necessary to give the meaning (semantic base) of the 5-line unit and to compare these to the words on the backing paper. Each student was further informed that one-fourth of a point would be awarded for each word correctly underlined; however, each extraneously underlined word would result in the loss of one-fortieth of a point from his or her daily score. This 1:4 ratio was determined from a previous series of studies (e.g., Zimmer *et al.*, 1978). We wanted students to earn points successfully, but we also wanted a reasonable response cost incurred for extraneous underlining. Students actually averaged 22.6 points toward their course grade with a range of 5.5 points to 42.4 points obtained. Students could earn up to 50 points. For example, if a person underlined 10 out of 14 correct words in one 5-line unit but also underlined 5 extraneous words, that person would receive .775 points for that portion of one essay. Scored for students as a part of the overnight scoring procedure, the results were reported during the following session (each person participating actually received the full 50 points toward the study skills unit regardless of scores).

Group assignments. The 30 participants were assigned to one of three groups at random. Group 1 received baseline (B) for sessions 1-5, the feedback (F) conditions for sessions 6-10, a return to baseline for sessions 11-15, the feedback plus points (F+) conditions for sessions 16-20, a return to baseline for sessions 21-25, and the feedback condition for sessions 26-30.

Therefore, the coded pattern for group 1 for the six time periods is B-F-B-F+-B-F. Group 2 received baseline for sessions 1-10, the feedback plus points conditions for sessions 11-15, a return to baseline for sessions 16-20, the feedback only condition for sessions 21-25, and a return to baseline for sessions 26-30. The coded pattern for group 2 is B-B-F+-B-F-B. Group 3 received baseline for days 1-15, the feedback only condition for days 16-20, a return to baseline for sessions 21-25, and the feedback plus points conditions for sessions 26-30. The coded pattern for group 3 for the six time periods is B-B-B-F-B-F+.

Response Measures

Essay posttests were used as an ongoing check on reading comprehension and were scored in the presence of the participants immediately after completion. For data analysis, results were expressed in percent correct form.

Student underlining responses were scored in two ways (Zimmer *et al.*, 1978; McCrudden, Note 2), "correct" underlining and "extraneous" underlining. Correct underlining consisted of the number of words underlined per essay that agreed with the semantic base. Extraneous underlining consisted of the number of words underlined per essay that did not agree with the semantic base. Both measures were scored by two independent raters not otherwise involved in the study, with results again expressed in percentages. That is, correct underlining was expressed as the per essay percentage of "meaningful" words underlined whereas extraneous underlining was expressed by using the number of correct words as a 100% base. Thus, if a student underlined 90 extraneous words and there were 40 correct words to be underlined, his or her percentage of extraneous underlining was 225%.

Reliability

Several different reliability checks were made throughout the study. Student on-task time was checked by the experimenters for each session

and cross-checked by an aide for 70 of the sessions. The mean of the absolute value of time disagreements was 4.2 sec for this sampling of 70 sessions. Pearson product-moment correlation reliability measures on the two measures were computed daily, with mean reliability ratings of .99 for correct underlining. In fact there were only four instances of disagreement on these response measures through the entire study.

Semantic Bases

The identification of the major ideas in the text, the semantic base, was accomplished by having two independent raters (cf. Glover, Zimmer, & Bruning, 1980; Zimmer et al., 1978) identify the fewest possible number of words in each 5-line unit which represented the "meaning" of those lines. For example, the five lines below:

In order for the pack to effectively function as a social unit there must be a minimal

amount of conflict between its members. This conflict is minimized through relatively complex patterns of dominance and submissive behaviors.

The semantic base for these lines was given as: "to function as social unit, minimal conflict present through dominance and submissive behaviors." The reliability of the raters' choices was determined for 5-line units by dividing the number of agreeing words by the number of agreeing plus disagreeing words. Reliability ranged from .70 to 1.00 for all the 5-line units with a mean of .88. The semantic bases determined by the raters' agreements were printed on the backing paper immediately underneath each 5-line unit on the label paper.

RESULTS

The results for correct underlining, extraneous underlining, and posttest results are presented separately below.

Fig. 1. Mean percent correct underlining.

Correct Underlining

Figure 1 represents the mean correct underlining scores for the three groups of students.

Days 1-5. All groups were in *baseline condition* during days 1-5, with virtually no differences among the mean scores of the three groups.

Days 6-10. *Group 1* was shifted to the *feedback* condition during days 6-10 while *groups 2 and 3* remained in *baseline* condition. Figure 1 pictures the sharp increase in correct underlining noted in group 1, from approximately 21% at day 6 to nearly 40% by day 10. No noticeable change in mean correct underlining was observed in groups 2 or 3 from days 6 to 10, with the mean for both groups remaining between 19 and 24%.

Days 11-15. During days 11-15, *group 1* returned to *baseline* conditions, *group 2* was placed under the *feedback plus points* condition, and *group 3* remained in the original *baseline* condition. Group 1 showed a slight decrease in correct underlining, averaging about 35.2% during re-imposition of baseline conditions. Group 2 showed a marked increase in correct underlining, similar to that observed from group 1 during days 6-10. Mean correct underlining for group 2 increased from approximately 21% to 36% by day 15, while group 3 remained at earlier baseline levels of approximately 22% correct.

Days 16-20. During days 16-20, *group 1* encountered the *feedback plus points* condition, *group 2* returned to *baseline* conditions, and *group 3* received the *feedback* condition. A sharp increase in correct underlining was noted for group 1, with correct underlining increasing from approximately 36% on day 16 to 53% on day 20. Scores obtained by group 2 stabilized at between 35 and 37% during return to baseline. Group 3, encountering an experimental condition for the first time, increased their mean underlining scores from 20% on day 15 to 40% on day 20, a change similar to that noted for group 1 during days 6-10 and group 2 during days 11-15.

Days 21-25. During days 21-25, *group 1* returned to *baseline* conditions, *group 2* received the *feedback* condition, and *group 3* returned to *baseline* conditions. Mean correct underlining scores for group 1 ranged from 50 to 52% during this period. Group 2 increased from approximately 41% on day 21 to 53% on day 25. Group 3 ranged from 36 to 48% during this period.

Days 26-30. During days 26-30, *group 1* received the *feedback* condition, *group 2* returned to *baseline* conditions, and *group 3* received the *feedback plus points* condition. As Figure 1 shows, the mean percent correct underlining scores for group 1 increased from 54% on day 26 to 60% on day 30. Group 2 evidenced a range of 52 to 53, while group 3 increased from 41 on day 26 to 55 on day 30, an increase similar to that observed during the second treatment imposition periods for both groups 1 and 2.

Statistical analyses. In order to examine the data more systematically, a repeated measures analysis of variance was computed over the results. Scores on days 5, 10, 15, 20, 25, and 30 were chosen as data points because they coincided with the final days of each treatment condition. Mean correct underlining was the dependent variable, with time and groups as independent variables. Significant differences in correct underlining were observed for both time, $F(5, 135) = 237.83, p < .01$ and groups, $F(2, 27) = 8.20, p < .01$. A significant groups by time interaction was also noted, $F(10, 135) = 7.72, p < .01$.

Because a logical result of the current design would be an interaction effect (groups received different treatments across times), the effects by groups were further analyzed with the Scheffé multiple comparison test. Results of the follow-up tests are presented in Table 1. No significant differences were observed among groups on the first (day 5) measure ($\bar{X}_1 = 20.0, \bar{X}_2 = 21.0, \bar{X}_3 = 22.0$). On measure two (day 10), group 1 (under the feedback condition) obtained significantly higher correct underlining scores

Table 1
Means and Standard Deviations by Group and Measure—Correct Underlining

Measure	Group 1 (B-F-B-F+-B-F)	Group 2 (B-B-F+-B-F-B)	Group 3 (B-B-B-F-B-F+)	Day Marginal Mean	F (between groups)
	$\bar{X}$ (SD)	$\bar{X}$ (SD)	$\bar{X}$ (SD)		
Day 5	Baseline 20.00 (7.27)	Baseline 21.00 (10.64)	Baseline 22.00 (9.78)	21.00	1.06
Day 10	Feedback 30.32 (3.53)	Baseline 21.00 (10.44)	Baseline 21.00 (7.63)	24.11	9.56*
Day 15	Baseline 35.10 (4.46)	Feedback 36.00 (7.77)	Baseline 20.00 (8.19)	30.37	8.72*
Day 20	Feedback+ 52.90 (7.04)	Baseline 36.22 (7.33)	Feedback 40.11 (4.33)	43.08	6.44*
Day 25	Baseline 51.12 (6.20)	Feedback 53.09 (5.27)	Baseline 37.64 (4.71)	47.28	5.12*
Day 30	Feedback 60.03 (4.65)	Baseline 53.42 (5.68)	Feedback+ 55.43 (3.32)	56.29	5.02*
Group Mar- ginal Mean	41.58	36.80	32.70		

*Significant at .01.

Note: Means joined by underlining not significantly different.

than both groups 2 and 3 ($\bar{X}_1 = 30.32$, $\bar{X}_2 = 21.0$, $\bar{X}_3 = 21.0$). No significant differences were noted between groups 2 and 3 (both still under the baseline condition). In comparing measures one and two, group 1 obtained significantly higher scores on measure two than on measure one. Groups 2 and 3 obtained scores that were not significantly different on measures one and two.

On measure three (day 15), no significant difference was noted between groups 1 and 2. Both, however, had significantly higher scores ($\bar{X}_1 = 35.1$, $\bar{X}_2 = 36.0$, $\bar{X}_3 = 20.0$) than group 3 (still under the baseline conditions). In contrasting scores on measures two and three, group 1 did not change significantly across time. Group 2 obtained significantly higher scores on measure three than on measure two. Group 3 did not change significantly between measures two and three.

On measure four (day 20), the scores obtained by group 1 were significantly greater than both groups 2 and 3, with no significant differences

between groups 2 and 3 ($\bar{X}_1 = 52.9$, $\bar{X}_2 = 36.2$, $\bar{X}_3 = 40.1$). When measures three and four were compared, group 1 showed a significant increase, group 2 showed no significant difference between the measures and group three showed a significant increase.

At measure five (day 25), the scores of groups 1 and 2 were not significantly different but both were significantly greater than the scores of group 3 ($\bar{X}_1 = 51.1$, $\bar{X}_2 = 53.1$, $\bar{X}_3 = 37.6$). When measures four and five were compared, group 1 showed no significant change across time. Group 2 demonstrated a significant increase in scores. Group 3 showed no significant change in scores between measures four and five.

On the last measure (day 30), the scores of group 1 were significantly greater than the scores of both groups 2 and 3 ($\bar{X}_1 = 60.3$, $\bar{X}_2 = 53.4$, $\bar{X}_3 = 55.4$). No significant difference was noted between groups 2 and 3. Comparisons of measures five and six indicated that both groups 1 and 3 showed significant increases whereas group 2 did not change significantly.

Fig. 2. Mean percent extraneous underlining.

Extraneous Underlining

Figure 2 presents mean extraneous underlining scores for the three experimental groups.

Days 1-5. All groups were in baseline conditions during days 1-5. No differences among groups were observed during this time.

Days 6-10. Group 1 (feedback) showed a marked decrease in extraneous underlining from nearly 190% on day 6 to 131% on day 10. Group 2 (baseline) generally followed the same pattern as during baseline with no evident change in responding. Group 3 (baseline) evidenced erratic changes in extraneous underlining varying from 205 to 154%.

Days 11-15. Group 1 (return to baseline) stabilized at near 130% with no apparent change. Group 3 (feedback plus points) dropped precipitously from nearly 200% on day 11 to 83% by day 15. Group 3 (baseline) evidenced a pattern very similar to that in the first days of baseline with the average percentage extra-

neous underlining varying from 184 to 204%.

Days 16-20. Group 1 (feedback and points) dropped from nearly 130% on day 16 to 46% on day 20. Group 2 (return to baseline) stabilized at 80 to 85% throughout the period. Group 3 (feedback) dropped from 198% on day 16 to 123% on day 20.

Days 21-25. Group 1 (return to baseline) ranged between 40 and 50% with no visible pattern of change. Group 2 (feedback) dropped from 82% on day 21 to 50% on day 25. Group 3 (return to baseline) ranged from 125 to 135% in a stable curve.

Days 26-30. Group 1 (feedback) showed a very slight decrease from 44% on day 16 to 37% on day 30. Group 2 (return to baseline) averaged 50% with little variation. Group 3 (feedback plus points) dropped from 132% on day 26 to 64% on day 30.

Statistical analyses. A repeated measures ANOVA was computed over the results using the same days employed for correct underlining as

Table 2
Means and Standard Deviations by Group and Measure—Extraneous Underlining

Measure	Group 1 (B-F-B-F+-B-F)		Group 2 (B-B-F+-B-F-B)		Group 3 (B-B-B-F-B-F+)		Day Marginal Mean	F (between groups)
	$\bar{X}$	(SD)	$\bar{X}$	(SD)	$\bar{X}$	(SD)		
Day 5	Baseline 201 (24.4)		Baseline 201 (25.9)		Baseline 203 (33.7)		201.67	.96
Day 10	Feedback 131 (11.8)		Baseline 200 (29.0)		Baseline 154 (38.2)		161.67	47.22*
Day 15	Baseline 128 (10.1)		Feedback + 83 (20.6)		Baseline 204 (25.1)		138.33	71.48*
Day 20	Feedback + 46 (9.8)		Baseline 85 (10.8)		Feedback 123 (11.1)		84.67	56.32*
Day 25	Baseline 46 (9.4)		Feedback 50 (4.4)		Baseline 127 (9.1)		73.33	59.22*
Day 30	Feedback 37 (8.2)		Baseline 50 (5.8)		Feedback + 64 (5.0)		50.01	10.92*
Treatment Marginal Mean	91.17		111.51		145.83			

*Significant at .01.

Note: Means joined by underlining not significantly different.

the measuring points. Significant differences were obtained between time, $F(5, 135) = 456.71$, $p < .01$ and between groups, $F(2, 27) = 57.01$, $p < .01$ as was a significant groups by time interaction, $F(10, 135) = 23.21$, $p < .01$.

Scheffé post hoc analyses explored the main effects as the interaction was an expected result given differing amounts of treatment in the different groups. The results of the follow-up tests are reported in Table 2. No significant differences were observed on the first measure. On measure two, group 1 (feedback) had significantly lower extraneous underlining scores than groups 2 and 3 ($\bar{X}_1 = 131$, $\bar{X}_2 = 200$, $\bar{X}_3 = 154$). Group 3 (baseline) was observed to have significantly lower scores than group 2 (baseline). In comparing measures one and two, groups 1 and 3 obtained significantly lower scores on measure two than measure one. No significant differences were observed between measures one and two for group 2.

On measure three, group 2 obtained signifi-

cantly lower scores than group 1 with both significantly lower than group 3 ($\bar{X}_1 = 128$, $\bar{X}_2 = 83$, $\bar{X}_3 = 204$). When time differences were examined, group 2 showed a significant decrease in extraneous underlining while group 3 showed a significant increase. No significant difference was noted between measures two and three for group 1.

On measure four, group 1 had significantly lower scores than group 2 with both significantly lower than group 3 ($\bar{X}_1 = 46$, $\bar{X}_2 = 85$, $\bar{X}_3 = 123$). Both groups 1 and 3 evidenced significant reductions in extraneous underlining between measures three and four. Group two did not change significantly between measures three and four.

On the fifth measure, no significant differences were observed between groups 1 and 2. Both, however, were significantly lower than group 3 ($\bar{X}_1 = 46$, $\bar{X}_2 = 50$, $\bar{X}_3 = 127$). Only group 2 evidenced a significant reduction in scores between measures four and five.

Fig. 3. Mean percent correct score on essay posttests.

On the last measure, a significant difference was observed between groups 1 and 3. Group 2, which had scores falling between groups 1 and 3, was not significantly different from either condition ($\bar{X}_1 = 37$, $\bar{X}_2 = 50$, $\bar{X}_3 = 64$). Only group 3 showed a significant change in scores between measures five and six.

Essay Posttest Results

Figure 3 presents the mean essay posttest scores.

Days 1-5. No differences were observed among the three groups (all in baseline).

Days 1-6. Group 1 (feedback) increased from 61% correct on day 6 to 73% correct on day 10. Groups 2 and 3 (both baseline) ranged from 54 to 64% correct with no noticeable trends.

Days 11-15. Group 1 (baseline) showed a flat curve with scores ranging from 72 to 73% correct. Group 2 (feedback plus points) increased from 62 to 74% correct between days 11 and

15. Group 3 (baseline) ranged from 56 to 59% correct between days 11 and 15.

Days 16-20. Group 1 (feedback plus points) changed from 74 to 79% correct during days 16-20. Group 2 (return to baseline) stabilized at about 72%. Group 3 (feedback) increased from 63 to 75%.

Days 21-25. Group 1 (return to baseline) was stabilized at 78 to 80% correct. Group 2 (feedback) increased from 70 to 80% correct. Group 3 (return to baseline) ranged from 73 to 75% correct.

Days 26-30. Group 1 (feedback) increased from 79 to 90 percent. Group 2 (return to baseline) ranged from 74 to 78 percent correct. Group 3 (feedback plus points) increased slightly from 73 to 80 percent correct.

Statistical analyses. A repeated measures ANOVA was computed over the results. Significant differences were observed between groups, $F(2, 27) = 3.20$, $p < .05$ and by time, $F(5,$

135) = 25.61, $p < .01$. There was no significant interaction.

In examining the between-groups main effect using a Scheffé follow-up on between-group mean difference, the group 1 posttest mean was significantly larger ($\alpha = .05$) than that for group 3. The group 1 to group 2 and group 2 to group 3 mean differences were not significant.

For the measures main effect, pairwise day mean comparisons were compared using a Scheffé multiple comparison procedure. Day 5 measures were uniformly significantly lower than those for the other day observations. Day 10 measures were significantly lower than those for day 20, day 25, and day 30, but not for day 15. Day 15 measures were consistently and significantly lower than for day 20, day 25, and day 30. Day 20 mean was significantly lower than that for day 30 but not for day 25. Day 25 mean was also not significantly lower than that of day 30. Results are summarized in Table 3.

Individual patterns. Individuals, overall, followed the same general pattern as the group

results. That is, correct underlining, extraneous underlining, and posttest scores remained fairly constant during baseline. When the first experimental condition was introduced, correct underlining increased, extraneous underlining decreased, and posttest performance improved. Returns to baseline did not result in extinction; instead, a flattening effect on the curves of all three response measures was observed. Reimposition of experimental conditions demonstrated further increases in correct underlining and posttest performance, while extraneous underlining was further decreased. An indication of individual patterns of performance may be seen in Tables 1, 2, and 3 which present the means and standard deviations for all three groups in the three response measures.

Reading Comprehension Pre- and Posttests

A t test for related samples assessed differences between pre- and posttest administrations of the comprehension section of the Nelson-Denny reading tests. A significant increase in

Table 3
Means and Standard Deviations by Group and Measure—Posttest Results

Measure	Group 1 (B-F-B-F+ -B-F)		Group 2 (B-B-F+ -B-F-B)		Group 3 (B-B-B-F-B-F+)		Day Marginal Mean	F
	$\bar{X}$	(SD)	$\bar{X}$	(SD)	$\bar{X}$	(SD)		
Day 5	Baseline 59 (5.85)		Baseline 59 (6.75)		Baseline 52 (9.79)		56.67	1.95
Day 10	Feedback 73 (4.53)		Baseline 62 (6.62)		Baseline 64 (6.10)		66.33	4.22*
Day 15	Baseline 72 (6.85)		Feedback+ 74 (6.47)		Baseline 58 (5.83)		68.00	4.26*
Day 20	Feedback+ 79 (5.42)		Baseline 71 (6.45)		Feedback 75 (5.19)		75.00	1.88
Day 25	Baseline 78 (6.54)		Feedback 80 (7.14)		Baseline 75 (5.03)		77.67	2.11
Day 30	Feedback 90 (7.93)		Baseline 76 (7.35)		Feedback+ 80 (4.03)		82.00	3.18*
Treatment Marginal Mean	75.17		70.33		67.33			

*Significant at .05.

Note: Means joined by underlining not significantly different.

comprehension was noted, $t = 32.11$, $p < .01$. Although the mean on the pretest was the 18th percentile with a range from the 10th to 24th percentile, the mean on the posttest was the 54th percentile with a range from the 48th to the 72nd percentile.

DISCUSSION

Within the design of the present study, it appears possible to modify operationally defined responses representing students' ability to identify the semantic base of prose passages. The correct underlining measures represent student attention to relevant sections of prose passages whereas the extraneous underlining measures represent student attention to nonsignificant portions of the essays. Both repeated measures analyses of variance and the Scheffé post hoc results strongly supported the graphically presented increases in correct, and decreases in extraneous, underlining. Additionally, essay posttest performance of students also increased, indicating an enhanced ability to recall and organize the topical information encountered.

The general procedures followed in the present study similarly increased student scores on standardized comprehension tests, demonstrating that these behavioral techniques may also be used to facilitate general comprehension strategies. Training students to reduce attention to insignificant parts of prose passages, thereby increasing comprehension of prose passages, would greatly enhance the effectiveness of existing reading instructional techniques.

The reader should note that neither the underlining measure nor the comprehension measure returned to baseline levels when the contingencies were removed. However, no changes in levels of responding were noted during baseline periods. Apparently changes in responding brought about by the experimental conditions were not conducive to extinction by merely removing feedback or feedback plus points. The varying lengths of baselines prior to treatment,

however, suggest that practice alone, or coupled with immediate feedback for essay posttests, did not bring about any change in participant responding. It appears clear that complex verbal behaviors can be brought under experimental control, and when operational definitions of educationally relevant behaviors such as attention and comprehension are developed, behavior analysis techniques may be used to enhance student processing of prose arrays.

An examination of the data seems to indicate there is no clear advantage of feedback plus points over feedback alone. In this study, it may be that points did not function as potent reinforcers or it may be that feedback is the important variable in enhancing measures of reading comprehension. As Zimmer *et al.* (1978) noted, feedback during the reading process is almost totally absent unless it is imposed by the reader himself. The provision of immediate feedback for identifying the semantic base of small segments of prose materials seems to have been the operative element in the current procedures.

The results of the current study should be cautiously interpreted. First, with our method of providing feedback, we altered connected discourse so that no segment of material was longer than five lines. Even though this was the case during baseline and treatment conditions, the prose materials we used did not exactly approximate "real" reading materials that students encounter. Second, instructions to identify the semantic base of text materials without feedback (e.g., Petersen *et al.*, 1980) have little effect on reading comprehension.

Despite these cautions, it appears that we should focus attention more carefully on techniques for analyzing complex verbal behaviors. Recent research in prose processing, semantic memory, and problem solving presents the opportunity to couple response measures and methodologies developed for these areas with the demonstrated strategies of applied behavior analysis to examine complex "cognitive" behaviors in greater detail.

REFERENCE NOTES

1. Glover, J. A., & Zimmer, J. W. *Nobody knows how to remember that prose*. Paper presented at the annual meeting of the American Educational Research Association, Boston, 1980.
2. McCrudden, T. *A contextualist analysis of adjunct aids in prose processing*. Unpublished doctoral dissertation. The University of Nebraska, 1979.
3. Zimmer, J. W. *A level of processing analysis of memory for prose*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978.
4. Zimmer, J. W. *A processing activities approach to memory for prose*. Paper presented at the annual meeting of the American Psychological Association, Toronto, 1978.

REFERENCES

- Ausubel, D. P. The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 1960, **51**, 267-272.
- Clawson, E. U., & Barnes, B. R. The effects of organizers on the learning of structured anthropology materials in the elementary grades. *Journal of Experimental Education*, 1973, **42**, 11-15.
- Faw, H. W., & Waller, T. G. Mathemagenic behaviors and efficiency in learning from prose materials: Review, critique and recommendations. *Review of Educational Research*, 1976, **46**, 691-720.
- Gagne, R. M. Analysis of objectives. In L. J. Briggs (Ed.), *Instructional design*, Englewood Cliffs, N.J.: Educational Technology Publications, 1977.
- Gagne, E. D., & Rothkopf, E. Z. Text organization and learning goals. *Journal of Educational Psychology*, 1975, **67**, 445-450.
- Glover, J. A., Ronning, R. R., Petersen, C. H., & Zimmer, J. W. A topographical analysis of prose processing activities. *The Journal of Genetic Psychology* (in press), 1980.
- Glover, J. A., Zimmer, J. W., & Bruning, R. B. An information processing approach to creativity: Memory from prose. *The Journal of Genetic Psychology* (in press), 1980.
- Hiller, J. H. Learning from prose text: Effects of readability level, inserted question difficulty, and individual differences. *Journal of Educational Psychology*, 1974, **66**, 202-211.
- Hunkins, F. P. Effects of analysis and evaluation questions on various levels of achievement. *Journal of Experimental Education*, 1969, **38**, 45-58.
- Jenkins, J. R., & Deno, S. L. Influence of knowledge and type of objectives on subject-matter learning. *Journal of Educational Psychology*, 1971, **62**, 67-70.
- Kintsch, W. Comprehension and memory of text. In W. F. Estes (Ed.), *Handbook of learning and cognitive process*, Vol. 6: *Linguistic functions in cognitive theory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.
- Nelson, M. S., & Denny, E. C. *The Nelson-Denny reading test*. Boston: Houghton Mifflin, 1973.
- Petersen, C. H., Glover, J. A., & Ronning, R. R. An examination of three prose learning strategies on reading comprehension. *Journal of General Psychology* (in press), 1980.
- Resnick, L. B., & Beck, I. L. Designing Instruction in Reading: Interaction of theory and practice. In J. T. Guthrie (Ed.), *Aspects of reading acquisition*. Baltimore: The Johns Hopkins Press, 1977.
- Rothkopf, E. Z. Learning from written instructive materials: An exploration of the control of inspection behavior by testlike events. *American Educational Research Journal*, 1966, **3**, 241-249.
- Smith, F. *Reading*. New York: Cambridge University Press, 1978.
- Zimmer, J. W., Petersen, C. H., Ronning, R. R., & Glover, J. A. The effects of adjunct aids in prose processing: A re-examination. *The Journal of Instructional Psychology*, 1978, **5**, 27-34.
- Zimmer, J. W., Glover, J. A., Ronning, R. R., & Petersen, C. H. On the utility of the Nelson-Denny Reading Test as a covariate in research on prose processing. *Perceptual and Motor Skills* (in press), 1979.

Received April 18, 1979

Final acceptance June 12, 1980