*ACADEMIC RESPONSES AND ATTITUDES ENGENDERED BY
A PROGRAMMED COURSE IN CHILD DEVELOPMENT¹*

GROVER J. WHITEHURST

STATE UNIVERSITY OF NEW YORK AT STONY BROOK

Both a small course section and a large course section were taught with methods that involved breaking material down into weekly units that had to be mastered by students as indicated by weekly tests. Attendance at lectures was voluntary and students were not tested on lecture material. In a system in which a student could receive only an A or an F for a course grade, less than 2% of the students receiving credit failed to master all material and received Fs. Students in the large course performed better on weekly tests and received fewer Fs than their counterparts in the small course, even though they rated their experience less favorably. Within the large course, the methods of tutorial interview, group discussion, and written assignment were compared in terms of their effectiveness in preparing students for weekly tests. Tutorials and written assignments were superior to group discussions in this regard. Students rated the effectiveness and enjoyability of these three methods in the reverse order from their actual effectiveness for test preparation. Some consequences of lack of congruity between attitude measures and performance measures were discussed.

Interest in the application of behavioral principles to university level teaching is burgeoning. Following Keller's (1968) lead, a variety of contingency management techniques have been applied to the teaching of experimental analysis principles (Ferster, 1968; Lloyd and Knutzen, 1969; Malott, 1971, Johnston and Pennypacker, 1971; Sheppard and MacDermot, 1970), as well as to the teaching of more traditional content (McMichael and Corey, 1969). Undoubtedly, other instructors have applied contingency management techniques to their classes without reporting the results of their efforts.

This report presents the results of efforts to control and analyze the academic behavior of undergraduates enrolled in a child development

course. A variety of purposes motivated the research. The primary technological goal was to engineer an educational environment that would be maximally effective in teaching small and large undergraduate course sections. As a subsection of this aim, three methods frequently employed as part of the instructional format in programmed courses were compared. These methods were tutorial interview, small group discussions, and written assignments.

Effectiveness of these techniques and overall effects of the course organization were analyzed using two different sets of responses: test performance measures and verbal evaluation measures. The measurement of evaluative reactions introduces possible outcomes that are not encountered when only performance measures are taken, as is the usual case with behavioral technologies. The criteria of verbal evaluation and performance may not always suggest the same conclusions about a particular educational design. What are the determinants of the two sets of response measures, and are they in any way orthogonal? As an outgrowth of the technological goals of the present study, preliminary an-

---

¹Much of the structure and content of the course described herein must be attributed to the model provided by Sidney W. Bijou, who has taught a similar course for a number of years at the University of Illinois. Appreciation is also expressed to Minna Goldfarb, Gary Novak, and Ross F. Vasta who contributed in many ways to the progress of this research. Reprints may be obtained from the author, Psychology Department, State University of New York, Stony Brook, Long Island, New York, 11790.

swers to these and other questions were sought. These questions go beyond technology to the issue of complex behavior, how it is to be characterized, and the variables of which it is a function.

METHOD

Subjects

In the Fall semester, 50 students enrolled in a section of a course entitled Child Development. This was one of three sections of the course offered that semester. Total enrollment in all Child Development sections for that semester was approximately 250. The following Spring semester, 247 students registered in the only section of Child Development to be offered. Students in the Fall had no way to anticipate before enrolling that the content or pedagogy of their course section would differ from that usually associated with the Child Development course, or from that being offered in the other two sections. In the Spring, many students were undoubtedly forewarned of the nature of the course, but the lack of alternative sections limited any selection biases that this might have introduced. In both the large and small course, approximately 65% of the students were psychology majors and most enrollees were juniors or seniors.

Procedure

Course content. The content of the course was arranged to (1) introduce the student to the basic principles of behavior analysis, (2) to apply these principles to a survey of data and theory in child development, and (3) to have the student recognize the application of the principles learned to applied problems in child-rearing and society. No single text in child development seemed to serve all these functions, and therefore reading was assigned from many sources. Students were required to read all or substantially all of two child development texts and a book of readings by Bijou and Baer (1961, 1965, 1967). They were assigned all of a pro-

grammed text by Geis, Stebbins, and Lundin (1965) and portions of a programmed text by Holland and Skinner (1961). Also required were portions of books by Skinner (1953) and Watson (1965). In addition to these materials, 11 journal articles were assigned and students also received five unpublished handouts covering topics such as schedules of reinforcement, basic and societal stages of development, and other theories of development. All this material constituted 811 pages of reading. Eight films were shown during the course of the semester. Formal lectures were delivered on seven occasions.

Small course organization. The content of the course section with the enrollment of 50 was organized into 13 units, each unit corresponding to one week of class time. The class met three times each week in hourly meetings. Students received a course syllabus on the first class meeting listing their assignments by units and the dates for various presentations in class.

Monday meetings of the class were devoted to films and formal lectures. Wednesday class meetings were devoted to group discussions of material assigned for that week. Students who arrived in class on Wednesday were divided into two or three groups, depending on the number of students present. One group remained in the meeting room and the other group(s) moved to public lounges in the same building. Individual group discussions were led by the instructor and two graduate teaching assistants assigned to the course. Group discussions were relatively unstructured. Each group discussion leader prepared his own list of questions to use in initiating dialogue within the group. The only instruction to the graduate student assistants was to get as many students talking about the material as possible. Group discussion leaders were free to answer any questions about the material raised by students during the course of the discussion.

Friday sessions were devoted to tests. The examinations consisted of 25 multiple choice items drawn solely from the reading material assigned for each unit. The tests were handed out at the

beginning of class and the student was instructed to mark his answers once on an answer sheet and once on the test. When he finished the test, his answer sheet was collected but he retained the test. When everyone had finished, correct answers were written on the blackboard and the remainder of the class time was spent discussing the examination and clarifying general misunderstandings.

The student was required to pass the multiple-choice test with three or fewer errors. If more errors were made, or if the student missed the test, he was required to attend a make-up involving an alternate form of the test on the Wednesday following the Friday test. This was conducted outside of regular class hours. The same error criterion was in effect. If the student failed two written tests, he was required to attend a one-to-one tutorial with one of the graduate assistants until the assistant was satisfied that the student had command of the material for that unit.

The final unit of the course was based on the theme of developing individuals interacting with a developing society. The usual weekly format was not followed. Discussions were held about experimental societies, educational institutions, and other societal-individual phenomena. Each student was required to submit a short paper in lieu of a test on this final unit. Papers were to be based on an application of the principles studied in the course to some societal or individual phenomenon. The papers were graded on a pass-fail basis. Unacceptable papers were returned to their authors with instructions to re-write them.

On the final meeting of the semester, a University sponsored questionnaire on reactions to the course was distributed and completed in class.

Students were informed at the beginning of the course that they could receive either an A or an F as a course grade. The grade of A would be given to all students passing all units by test or tutorial and submitting an acceptable paper before the end of the semester. Any student who did not complete all units and submit an accept-

able paper would receive an F. Students could also receive withdrawal grades (no-credit) in ways consistent with university regulations.

Large course organization. The course with enrollment of 247 students was conducted in a large lecture hall with approximately 450 seats. In addition, the university provided a large office for use by course assistants and an additional lecture hall on Wednesdays. One graduate student assistant was added to the two assistants who had been assigned to the small course. Ten undergraduates who had received the grade of A in the small course volunteered to serve for academic credit as tutors in the large course. They were told that their grade would be based on their completion of all assigned duties and on the performance of their students.

The Monday course meetings remained the same as they had been in the small course. Lectures were given and films were shown. Students were not directly tested on these films and lectures and were not required to attend.

Friday meetings also remained the same. The same tests were employed. The same error criterion was in effect.

Make-up procedures were also the same, with the following exception. If a student failed a multiple-choice make-up test, he had to submit answers to a prepared set of questions on his next attempt to pass that particular unit. The interview with a graduate assistant was reserved for those students who had failed a multiple choice make-up test and who had failed the long answer written questions.

The significant changes from the procedures used for the small course were made for the Wednesday class meetings. For the small course, unstructured voluntary group discussions were held. For the large course, students were given at the beginning of each unit a list of discussion questions. There were usually eight or nine detailed questions. Some questions were designed to make the students integrate written material assigned for the unit; other questions were designed to focus the students' study on relevant portions of the reading; and other questions

were aimed at points in the reading or on the tests that had been particularly confusing to students in the small course. The discussion questions covered enough of the assignment so that preparation of the answers after having read the material should have ensured a passing mark on the Friday test.

Students were required to participate in six Wednesday exercises based on these discussion questions during the course of the semester. There were three different types of Wednesday exercises available. Students could hand in written answers to the questions. Written corrections to these answers were supplied by their undergraduate tutor and the student could pick up his corrected answers on Thursday morning. The second type of exercise involved the student engaging in a one-to-one tutorial with his undergraduate tutor. These tutorials were held during the Wednesday class meeting. They were scheduled to last 10 min. The tutors had previously been supplied with correct written answers to the discussion questions. The tutor would ask the student to answer each of the questions from the question list in turn. After the student provided an answer, the tutor would correct misinformation. If the student came to his tutorial without having prepared answers, the tutors were instructed not to provide answers for the student and to terminate the tutorial. Finally, students could participate in a group discussion based on the discussion questions. Four group discussions were held each Wednesday, led by the three graduate student assistants and the instructor. Group discussions lasted for an hour. The group discussion leaders were instructed to go through the questions eliciting answers from students. Misunderstandings and incorrect answers were to be corrected and related topics were to be pursued as time permitted. Group discussions were held simultaneously in different locations in the main lecture hall and the smaller lecture hall that had been provided.

Students were required to participate in two of each of these three types of Wednesday exercises, making six required exercises in all. At the

beginning of the semester, each student was given information indicating those Wednesdays when he was required to participate, the method of participation required on a particular Wednesday, and the location in the lecture halls to which he was to report. These requirements were so organized that a particular student had to participate every other week over the first 12 units. The order of participation in the three types of Wednesday exercises was counterbalanced across the six exercises for each student. Students were told that on the Wednesdays when they were not required to participate, they could participate in the Wednesday exercise of their choice or they could do nothing.

The result of this organization for the students as a whole was that half the students were required to participate in a Wednesday exercise one week, and the second half the next week, and so forth through the semester. In a given week, one-third of the students who were required to participate handed in written answers; one-third went to tutorials; and one-third went to group discussions. All possible counterbalanced orders of Wednesday exercises were represented across students. Assignment to a particular order was random.

After the eighth unit of the course, these procedures were changed so that attendance at group discussions became completely optional; attendance at tutorials and written assignments was still required. The nature of the group discussions was also altered. Group discussions no longer were focused on the discussion questions but were allowed to proceed in an unstructured manner, as they had in the small course.

At the final class meeting, questionnaires about reactions to the course were distributed and completed.

RESULTS

Both the small and the large course produced a large proportion of students receiving grades of A. In the small course, 92% of the students appearing on the official course roster and tak-

ing more than two tests received the grade of A; 6% of this group of students withdrew from the course and received no credit; 2% of these students received the grade of F. In the large course, 96% of the students appearing on the official course roster and taking more than two tests received grades of A, 3% withdrew, and 1% received Fs. Evidence related to the question of whether these grades were fairly earned comes from responses to the anonymous questionnaire given at the end of the small course. Eighty-two per cent of the respondents indicated that they had done all of the required reading in the course and the remaining 18% indicated that they had done most of the work.

Further comparison of the large and small course can be made on the proportion of students failing their first test and having to take some form of make-up examination. In the small course, an average of 14% of the students receiving a final letter grade failed the first form of their unit test in a given week. This figure was 12% for the large course.

An administrative failure on the part of the university made it impossible to use the same attitude survey for both the large and the small course. Nevertheless, some data on the comparative student evaluation of the two courses can be obtained by examining responses to a similar question occurring on both survey forms. Students in the small course were asked to rate their instructor on a five-point scale, compared with all other instructors they had had at the university. The mid-point of this scale was labelled "about average". Students in the large course were asked to rate their course on a seven-point scale compared with all other courses they had had at the university. The mid-point of this scale was labelled "about the same". Of the 38 respondents in the small course, none responded below the mid-point of the scale; in the large course, 23% of the 181 respondents rated the course below the mid-point of the scale. Further comparative data on the effectiveness of the two courses in generating interest and positive evaluations can be obtained by comparing attend-

ance figures for the voluntary Monday class meetings. In the small course, attendance on Mondays dropped at its lowest point to 30% of the enrolled students. In the large course, attendance at its lowest point dropped to 8% of the enrolled students and after the first two units was never higher than 25% of the enrolled students.

Within the large course format, it is possible to compare the effectiveness of attending a tutorial, attending a group discussion, handing in written answers, or doing nothing on Wednesday in terms of unit test scores. Since the procedures of the course were changed after Unit 8 to make group discussions voluntary, data for the analysis were drawn from the first eight unit tests only. Also, since some students through Unit 8 had not engaged in all of the three types of Wednesday exercises, data were entered in the analysis from only those students who had participated in all three types of Wednesday exercises and who had done nothing on at least one Wednesday through Unit 8. A single factor repeated measures analysis of variance (Winer, 1962) was conducted on these data ($F = 4.17$, $df = 3, 612$, $P = < 0.01$), and the hypothesis that there was no differential effect on unit tests attributable to the type of Wednesday exercise engaged in was rejected.

The mean number of incorrect answers on multiple choice tests after preparing with written exercises was 1.56, for tutorials 1.65, for group preparation 1.88, and for *no* formal preparation, the mean number of incorrect answers was 1.90. A comparison of the average of the tutorial and written means with the average of the group and nothing means was significant at the 0.01 level. None of the remaining orthogonal comparisons were significant. Thus, answering discussion questions in a tutorial or in a written form resulted in significantly fewer errors on the unit tests than did attending a group discussion or engaging in no formal activity in relation to discussion questions. The first two methods did not differ significantly from each other, nor did the latter two methods.

Contrasted with these results obtained using a performance criterion are the results from the attitude questionnaire administered in the large course. Replying to a question about which type of Wednesday exercise they found most helpful in preparing for the Friday test, 44% of the students picked the group discussion method, 34% picked the tutorial, and 22% picked the written method. Replying to a question about which type of Wednesday exercise they found most enjoyable disregarding preparation for the Friday tests, 62% chose the group discussion, 33% chose the tutorial method, and the remaining 5% chose the written exercise.

The effect of changing the Wednesday group discussions from required to voluntary, while also changing the nature of those discussions, was assessed using the measure of class attendance. Class attendance per week for the three

Wednesday exercises is presented in Figure 1. Note the dramatic drop in attendance for the group discussions subsequent to the change of procedures after Unit 8 compared with the base-lines for the other exercises.

DISCUSSION

Evaluated against test performance criterion, both the large and the small course were highly successful. The large course was slightly more successful than the small course in this regard with more students passing the first forms of unit tests and more students getting As. If one assumes that a dramatic increase in class size should result in poorer student performance, then the equality or superiority of the large course must be attributed to the introduction of

Fig. 1. Number of students performing each of the three types of Wednesday exercises by course unit in the large course. After Unit 8, attendance at the group exercise was made optional and discussions in groups no longer focused on discussion questions.

study questions and associated review procedures in the large course.

While not evaluated experimentally, it is quite probable that the high proportion of As obtained by both classes was due in part to the provision of only one other alternative to an A, an F. Students who would otherwise have been satisfied with a grade lower than an A performed at high levels in order to avoid an F. The desirability of such a procedure might be questioned if evaluated against standards other than test performance. Students working in an avoidance paradigm for an A may be less likely to use the content of the course in new situations; they may dislike the course and by verbalizing their dislike, harm interactions between instructors and other students; they may be more likely to cheat and engage in other "grade-oriented" behaviors. These questions remain to be evaluated. The A-F procedure also deprives the university of the numerous grades differentials it requires in order to confer distinctions upon students. This may or may not be a negative attribute of the procedure.

Despite the equality or superiority of the large course on test performance criterion, there is the evidence from attitude surveys and from voluntary class attendance that the large course was liked less than the small course and generated less student interest and involvement. These data are strongly supported by the impressions of the instructor and graduate assistants who were struck by the unfavorable change in the nature of interactions with students that occurred with the change to the large class. Student-staff relationships were often adversarial in the large course but there was an almost complete absence of this type of interaction in the small course. It may well be that when class size reaches the level at which the instructor cannot recognize his students when he sees them outside of class and at which most of the students feel no personal contact with the student in the adjacent seat, there are changes in the normal controlling functions of social stimuli. Students for whom the approval of peers or the instructor is a

powerful reinforcer may react very differently to these same stimuli when delivered by peers and an instructor who do not know the students. There is a strong suggestion that these effects occurred in the large class, notwithstanding the fact that students were learning more according to performance criteria. Such effects should not be ignored by those planning the policies of instructional institutions, for they probably influence to a high degree collateral response measures such as student interest and involvement, which may be more predictive of ultimate educational objectives than are course grades.

While the three types of Wednesday exercises did not produce large absolute differences in performance on the Friday tests, the tests were so designed that they produced an extremely constricted range of scores. Thus, the differences that the separate preparation methods would produce on a more discriminating measure would probably be substantially larger. The relative effectiveness of the three types of Wednesday exercises in preparing students for Friday tests was reversed from informal predictions made by the instructor before instituting the procedures. It was reasoned that since the instructor and the graduate students had much more information at their disposal than the undergraduate tutors, the students would learn more from them than from the relatively unsophisticated undergraduates. An explanation for the obtained results may be as follows: a student having to prepare for a tutorial or having to hand in written answers was forced to read the material in order to prepare his answers. Students coming to a group discussion, in contrast, could do relatively little preparation and obtain answers to the discussion questions from the few well-prepared students as they volunteered responses to the group leader. Armed with these pilfered answers to the discussion questions, students could do reasonably well on unit tests but could not anticipate subtleties in questions that were obvious only after having read the assigned material. Thus, these students did relatively less well on unit tests.

One function of the change in procedures after Unit 8 was to test the preceding hypothesis. The drop in attendance as the group leaders no longer focused on the answers to discussion questions could of course have been a result of either the change in requirements or the change in the nature of the discussion section. However, the fact that 28 students came to the discussion group for Unit 9, even though it was completely voluntary, coupled with the gradual drop in attendance thereafter, suggests that students no longer attended when they found that they could not obtain full answers for their attendance.

The disparity between the attitude measures of the Wednesday exercises and the test performance measures of these exercises is interesting. Note that one is the mirror image of the other. Presumably, the students' evaluation of the enjoyableness of the three methods influenced them to judge incorrectly the effectiveness of the methods in terms of test performance. There are numerous related findings from the fields of perception (Goldiamond, 1958) and social psychology (Kutner, Wilkins, and Yarrow, 1952) indicating that subjects' verbal reports are not always highly correlated with other response measures. Confronted with these kinds of data, there has been a tendency, particularly among "operant" psychologists, to resolve the dilemma by rejecting verbalizations and attitude statements as less-reliable indicators of environmental effects than non-verbal measures. As translated into a classroom analysis, this would mean that instructional effectiveness would be defined in terms of test performance, not attitude statements. This may be unfortunate. On intuitive grounds, one would predict that the more students enjoy an educational experience, the more likely they are to pursue the content of that experience beyond the bounds of the classroom. This is clearly desirable. On the other hand, if students do not acquire certain skills as a result of exposure to a technology, they will be poorly equipped to advance their educations in an independent fashion or to make immediate

use of the material that has been presented to them.

This analysis suggests that we need to know many more of the determinants of what seem to be partially independent measures. We need to develop technologies that produce skill acquisition and student satisfaction. The practice of deriving measures of teaching success from only one of these areas is likely to produce instructional techniques that are much less effective than they could be in producing long-term learning and involvement.

REFERENCES

- Bijou, S. W. and Baer, D. M. *Child development*, Vol. 1. *A systematic and empirical theory*. New York: Appleton-Century-Crofts, 1961.
- Bijou, S. W. and Baer, D. M. *Child development*, Vol. 2. *Universal stage of infancy*. New York: Appleton-Century-Crofts, 1965.
- Bijou, S. W. and Baer, D. M. *Child development: readings in experimental analysis*. New York: Appleton-Century-Crofts, 1967.
- Ferster, C. B. Individualized instruction in a large introductory psychology course. *Psychological Record*, 1968, 18, 521-532.
- Geis, G. L., Stebbins, W. C., and Lundin, R. W. *Reflexes and conditioned reflexes*, Vol. 1, Part 1. New York: Appleton-Century-Crofts, 1965.
- Goldiamond, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: analysis in terms of psycho-physical indicator methodology. *Psychological Bulletin*, 1958, 55, 373-411.
- Holland, J. G. and Skinner, B. F. *The analysis of behavior*. New York: McGraw-Hill, 1961.
- Johnston, J. M. and Pennypacker, H. S. A behavioral approach to college teaching. *American Psychologist*, 1971, 26, 219-244.
- Keller, F. S. "Good-bye, teacher . . ." *Journal of Applied Behavior Analysis*, 1968, 1, 79-89.
- Kutner, B., Wilkins, C., and Yarrow, P. R. Verbal attitudes and overt behavior involving racial prejudice. *Journal of Abnormal and Social Psychology*, 1952, 47, 647-652.
- Lloyd, K. E. and Knutzen, N. J. A self-paced programmed undergraduate course in the experimental analysis of behavior. *Journal of Applied Behavior Analysis*, 1969, 2, 125-133.
- Malott, R. W. *Contingency management in an introductory psychology course for 1000 students*. Kalamazoo: Behaviordelia, 1971.
- McMichael, J. W. and Corey, J. R. Contingency management in an introductory psychology course

- produces better learning. *Journal of Applied Behavior Analysis*, 1969, 2, 79-83.
- Sheppard, W. C. and MacDermot, H. G. Design and evaluation of a programmed course in introductory psychology. *Journal of Applied Behavior Analysis*, 1970, 3, 5-11.
- Skinner, B. F. *Science and human behavior*. New York: MacMillan, 1953.
- Watson, R. I. *Psychology of the child*. New York: Wiley, 1965.
- Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- Received 24 September 1971.
(Revised 10 January 1972.)