A STUDY OF ERROR-MODELING IN SKILLS LEARNING

By NORA I. PARKER*
Professor of Nursing
University of Toronto

INFORMATION feedback, the process of providing information about the correctness or adequacy of a response (Bilodeau, 1966) (1), is regarded as the single most important variable governing the acquisition of skilled habits (Bilodeau, 1966; Irion, 1966). Fitts (1965) proposes a model which distinguishes between internal and environmental feedback loops. In education and/or training, comments from an instructor would be one source of environmental feedback. In addition, in many skills learning tasks, feedback intrinsic to the task would also form part of the information coming from the environment (Irion, 1966). However, in other tasks, clarity of feedback from the task itself is minimal and information from the instructor then assumes greater importance.

Markle (1965) proposes, in regard to the teaching of English, that it may be possible to teach the student appropriate discriminations so that he may become his own observer and evaluator, and, with this self-monitoring behavior, be able to provide his own information and knowledge of results rather than having to rely on environmental feedback. This would be of particular value where feedback intrinsic to the task is low or non-existent.

To enable the student to make such discriminations between correct and incorrect responses in his own repertoire, it would appear that prior information about errors as well as about correct responses would be helpful. This would be consistent with Olson's (1971) view that any form of instructional system is successful to the extent that it provides the learner with an awareness of critical alternatives and how to choose among them.

The view that learners should be supplied with information about errors is, of course, contrary to the generally accepted approach to instruction which emphasizes minimization of errors. In the case of skills learning particularly emphasis on teaching has been on correct performance usually modeled by an expert. In some instances however, the question of whether it would facilitate learning to demonstrate negative as well as positive instances of the behavior in

---

* This study formed part of a doctoral thesis completed at the University of Toronto. The author wishes to acknowledge the help of Students and faculty who participated in the project.

question has been raised (Allen and Ryan, 1969; Hunt, 1971; McDonald and Allen, 1967). The use or value of errors in educational practice has, however, been inadequately investigated and the purpose of this experiment was to investigate the possible value of modeling negative instances of certain actions involved in a complex procedure in nursing. The preparation of a surgical dressing tray was selected as the procedure.

The experimental hypotheses were based on the following propositions derived from theories of skills learning:

1. That cognitive processes utilizing information underlie the performance of a motor skill.
2. That understanding usually achieved through exposition and positive exemplary modeling would be enhanced by information about common errors.
3. That this will lead to enhanced performance of the skill as it enables the learner to choose between correct and incorrect alternatives.
4. That stored information regarding correct and incorrect alternatives is used to assess the performance of others as well as to direct one's own behavior.

METHOD

Experimental Design

The experimental design was a two factor design with repeated measures on one factor (practice), and called for three groups of subjects. It was decided that two groups of subjects would be exposed to the experimental condition (error modeling) and the third group constituted a control. With A as treatment and B as practice the design can be represented as follows:

Treatment	Subjects	Trials	
		B_1	B_2
A_1 (Errors A & B)	23	A_1B_1	A_1B_2
A_2 (Errors C & D)	24	A_2B_1	A_2B_2
A_3 (Control)	24	A_3B_1	A_3B_2

Subjects

Subjects were 71 second year Basic Degree students at the University of Toronto Faculty of Nursing. Subjects were all female with an average age of 20 years 1 month. Due to limitations of time, no pre-testing of the subjects in the skill was carried out. It was felt that random assignment to experimental and control groups gave some assurance that any differences in the control and experimental groups would be unbiased with respect to the experimental hypotheses (Campbell and Stanley, 1963). Prior to exposure to the experimental conditions, all subjects received the same preparatory teaching in relation to asepsis.

Modeling Device (2)

Three video tapes modeling the setting up of a dressing tray were produced. The demonstration variable (Sheffield and Maccoby, 1961) built into the tapes involved the modeling of errors which had been found to occur frequently both in the preparation and application of surgical dressings. Each of these errors were due to the violation of some principle or principles of surgical asepsis and results in contamination of hands or equipment.

The unit of the task was modeled in the same way on all three tapes except for the following variations:

1. Tape 1 — Standard procedure with no errors;
2. Tape 2 — Standard procedure except for the inclusion of two errors A and B;
3. Tape 3 — Standard procedure except for the inclusion of two other errors — C and D.

Video Tape for Discrimination Test (3)

A video tape showing a complete dressing procedure was produced. This tape contains 14 errors which include the four which were incorporated in the modeling tapes. The errors in this tape are executed in a sophisticated manner and, in contrast to those in the modeling tapes, most are not easily detectable except by someone skilled in aseptic technique. The use of this tape is described in the procedure.

Procedure

During the period of preparatory teaching prior to the project being carried out, the experimenter met with the total group of students to explain the plans for the research and request their cooperation.

Modeling — Following the preparatory teaching the modeling tapes were shown to the three groups with methods of utilization held as constant as possible among groups.

Viewing of the tapes was followed by a discussion with each group. Because of the possibility of introducing experimenter bias, the discussions were led by a person who had no other involvement with the research. Guidelines for the discussion were prepared by the experimenter. Briefly, the subjects in the experimental groups were asked to identify the modeled errors, relate them to the principles involved and discuss why they are considered to be errors of application. The discussion in the control group concentrated on positive exemplars of the same principles.

Practice Sessions — These began the day after the modeling tapes were shown. Each subject had two practice sessions or trials during which she was observed as she proceeded through the preparation unit of a surgical dressing procedure by one of six observers.

Six members of the University of Toronto Faculty of Nursing participated in the project as observers. All had had experience teaching and evaluating student performance in the clinical laboratory. An analysis of variance was used to estimate observer reliability. The rationale for this method is discussed by Winer (1970) but briefly is as follows. The procedure is based on the idea that in every measurement there is a true magnitude of whatever is being measured plus error of measurement. Upon repeated measurements of the same object with comparable instruments (in this case, the six observers) it is assumed that the true magnitude remains constant while the error of measurement varies. If the true measurement remains constant the variance within is due to error of measurement and the pooled within person variance is used to estimate this error. With a number of items being measured (in this case, the seven students) the variance among will also be partly due to differences in the true magniture. The ratio of the variance of effects due to true measurement to the error variance is used to arrive at the reliability coefficient. The estimate of inter-observer reliability from the among and within-subject variance was .949.

In both practice periods, subjects proceeded through the task without interruption for instruction, correction or feedback of any type. Observation was limited to noting the number of times any of the four modeled errors were made by each subject during each practice period. If the same error was made more than once by a subject, it was counted each time it occurred. The number of times the four errors occurred constituted the dependent variable. No other aspects of performance were taken into account.

Discrimination Test — Based on the proposition that stored information about incorrect as well as correct alternatives is used in assessing the performance of others, a test to determine whether the error modeling had increased skill in discriminating errors made by another person was set up.

In this test, given one week after the final practice, the students in both the experimental and control groups viewed the video tape of the complete procedure which has previously been described. The ability of subjects to detect the errors in this video constituted the test of discrimination. Since for each experimental group there were 12 errors to which they had not previously been exposed, it was possible to determine whether the ability to discriminate would extend beyond the two to which the experimental subjects had been sensitized.

In the instructions to the subjects, no mention was made of errors. They were instructed that they would be asked to evaluate the performance of the nurse in the video and to do this would be asked to complete a "rating scale". Twenty-five items were included on the scale and for each, subjects were asked to give a rating and then to justify the rating they gave. The ratings were actually not used in any way. Rather, the answers given to justify the rating were used to determine whether the subject had noted the error. Thirteen of the items on the scale were such that if the error had been noted, it would be evident in the answer. Returns were coded and were scored independently according to predetermined criteria by another person and by the experimenter. Any discrepancies were resolved while returns were still coded.

RESULTS

Error Scores — Practice Periods

The error scores from the practice periods were first subjected to statistical analysesaccording to the original design. The analysis of variance which tested a treatment effect, trials effect and treatment X trials interaction is shown in Table 1.

Table 1
Analyses of Variance of Error Scores

Source	Degrees of Freedom	Mean Square	F test	Probability
Between	70			
Treatment	2	25.590	.530	N.S.
Error	68	48.275	--	—
Within	71			
Trials	1	298.845	9.804	.005
T x T	2	61.785	2.027	N.S.
Error	68	30.479	—	—

Table 2
Analyses of Variance of the Total Number of Errors Discriminated

Source	S.S.	D.F.	M.S.	F	P.
Among	14.377	2	7.188	4.692	.01
Within	101.100	66	1.531	—	--
Total	115.477	68		—	—

The results showed only a significant trials effect. The modeling condition did not produce any differences and the treatment X trials interaction did not reach a significant level.

A chi-square analyses of the number of subjects who met a criteria of adequate performance on each of the separate errors was also carried out but none of the chi-square values obtained reached a significant level.

Although there were no reasonable grounds on which to support the idea that modeling errors will decrease their frequency in practice, there was, on the other hand, no evidence that modeling errors increased the frequency of their occurrence.

Discrimination test

A simple analyses of variance was carried out on the total number of errors discriminated by each subject. The analyses appears in Table 2.

Since the F was found to be significant, Duncan's Multiple — Range test (Edwards, 1960) was used to determine which of the means differed significantly from each other. Summary of the comparisons appears in Table 3.

The test showed that the means of both experimental groups differed significantly from the control group mean (p. 05) but not from each other.

Similar analyses were carried out on the unmodeled errors only, and on the modeled errors alone.

The results of the discrimination test, which with one exception were in accord with the predicted results, can be summarized as follows:

1. A significantly larger number of errors were discriminated by the subjects in both experimental groups than by the control subjects.
2. Group 1 subjects were able to detect a significantly larger number of the errors that had been modeled for them than either of the other two groups.
3. Group 2 subjects detected more of the errors which had been modeled for them than did the control group subjects but not a significantly larger number than the other experimental group.
4. In an analysis of unmodeled errors only Group 2 subjects exceeded the control subjects by a significant degree.

DISCUSSION

This study is subject to the limitations inherent in experiments under classroom conditions where numerous variables have to be taken into account. Control of variables known or thought possibly

Table 3

Multiple Comparisons Among Means of All Errors

Means	Control	Exp. Gp 1	Exp. Gp 2	Shortest Sig. Range (P. 05)
	4.29	5.18	5.30	
4.29	—	.89	1.01	$R_2 = .718$
5.18	—	—	.12	$R_3 = .755$

to affect learning, such as rote versus meaningful material, expository versus inductive teaching and verbalization by students of principles was taken into consideration since the teaching methods other than the experimental conditions imposed were the same for all students as was the content. Practice periods had to be spaced over a period of days because of limitations of space and observers. This was controlled by allotting subjects from each of the three groups to each practice period.

The results of the initial part of this experiment did not provide support for the notion that developing an awareness of common errors will enable the learner to avoid them in choosing among alternative actions and thus improve performance on a skill. On the other hand, contrary to usually accepted points of view there was no negative transfer effect (due to the modeling of errors) noted.

The discrimination test on the other hand, did provide some evidence in support of the proposition that stored information regarding incorrect alternatives does lead to more acute discrimination than does knowledge of the correct procedure only.

The absence of any statistical difference in the performance of the three groups on the practice trials, raises the question of why the error modeling did affect performance on the discrimination test but not on the performance of the perceptual-motor task. Several alternatives suggest themselves.

First of all there is a question of the adequacy of the information which was supplied. The increment in information provided by the error modeling consisted only of information as to what errors commonly occur. Neither the modeling nor the discussions included instruction as to how to avoid errors or gradually attain competence (Meichenbaum, 1971). It appears that there might be value in adapting Meichenbaum's "coping model" to educational purposes and that such use should be tested empirically.

Another possible reason that no differences were found in the practice trials relates to the transfer of training required in the two tasks. The perceptual-motor skill required a more remote type of transfer of learning than did the discrimination task which involved perceptual-symbolic responses similar to those called for under the error modeling conditions. In discussing the training of teachers, Hunt (1971) distinguishes between skill in discrimination and performance skill. This may result in the trainee being unable to perform the more complex task even though he can discriminate what is necessary. Because of this distinction, Hunt suggests that training in both discrimination skill and performance skill is necessary, the first being prerequisite to the second. Further, he suggests that one reason for failing at the more complex task, that is, the performance, may be due to inadequate training in discrimination. The distinction made by Hunt (1971) may be a useful one for nurse educators to consider and skill in discrimination in many situations would be a worthwhile educational objective in itself.

Apart from the statistical results of the study, certain observations made during the course of this investigation might also prove of interest to educators. These relate to certain assumptions regarding 'negative instances', 'correct procedures' and continuous corrective feedback.

With respect to the use of negative instances or errors in education, it was noted that, despite expressed concern about using these, they are used frequently for illustrating points to students. This became apparent when material for preliminary teaching and for the discussion groups was reviewed by the experimenter. It seems that, rather than positive instances of the application of principles, illustrations often involved errors. Does this include a prescription of 'Don't do this'? In any case, it seems that there should be an awareness of the frequent use of negative instances for illustrative purposes.

It was also noted that the instructors who participated as observers as well as many of the students, found a situation in which continuous instruction and/or immediate corrective feedback is *not* provided, uncomfortable. The goal of fixating 'correct' procedures is probably closely related to circumstances in the hospital clinical laboratory; however, one can speculate as to how students develop ability to monitor their own performance if they are receiving continuous instruction or correction. Under these circumstances, how and when does the student learn to function effectively on her own?

One last observation but one which appears critical, concerns the difficulties which were encountered when the observers with the experimenter tried to define what behavior constituted an error. After considerable discussion, agreement was reached *for the purposes of the study*. There was considerable difference of opinion as to when an action was in fact an error and when it was simply an alternate way of applying a principle. Since the contingencies are unobservable, the answer is difficult, but one was left wondering how much was essential and how much was rationalization of a ritual. One could conceive of ways of determining whether an action had actually resulted in contamination of an area. The differences with respect to what would be considered acceptable by one instructor yet not by another, also seems to have some implications for evaluation of student performance.

Notes

1. Bilodeau differentiates between the process of providing information during the course or at the end of a response and the subject's processing of that information. She suggests that 'Knowledge of results' be used to refer to the latter.
2. The term 'modeling' is being used here in a broader sense than in psychological research, where it is the exact reproduction of a response. It is here being used for demonstration of a behavior that illustrates a category of responses all ending in the same contingency-contamination.
3. This video tape has since proven useful as a teaching device, as it stimulates discussion by students of the application of principles and alternate ways to apply principles.

References

Allen, D. and Ryan, K. *Microteaching*. Don Mills, Ontario: Addison-Wesley, 1969.

Bilodeau, I. M. Information feedback. In E. A. Bilodeau (Ed.), *Acquisition of skill*. New York: Academic Press, 1966.

Campbell, D. T. and Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally and Co., 1966.

Edwards, A. L. *Experimental design in psychological research*. (Rev. ed.) New York: Rinehart and Co., 1966.

Fitts, P. M. Factors in complex skill training. In R. Glasser (Ed.) *Training research and education*. New York: John Wiley, 1965.

Hunt, D. E. *Matching models in education: The coordination of teaching methods with student characteristics*. Toronto: O.I.S.E. Monograph Series No. 10, 1971.

Irion, A. L. A brief history of research in the acquisition of skill. In E. A. Bilodeau (Ed.) *Acquisition of skill*. New York: Academic Press, 1966.

Markle, S. M. Programmed instruction in English. In R. Glazer (Ed.), *Teaching machines and programmed learning, II: Data and directions*. Washington: National Education Association, 1965.

McDonald, F. J. and Allen, D. W. *Training effects of feedback and modeling procedures on teaching performance*. Stanford University, 1967.

Meichenbaum, D. H. Examination of model characteristics in reducing avoidance behavior. *Journal of Personality and Social Psychology, 1971, 17. 298-307*.

Olson, D. R. *On a theory of instruction: Why different forms of instruction result in similar knowledge*. Paper presented to the International Symposium on the Problems of Suggestology, Varna, Bulgaria, June, 1971.

Sheffield, F. D. and Maccoby, N. Summary and interpretation of research on organizational principles in constructing filmed demonstrations. In A.A. Lumsdaine (Ed.), *Student response in programmed instructions: A symposium on experimental studies of cue and response factors in group and individual learning from instructional media.* Washington: National Academy of Science, 1961.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.