*THE EFFECTS OF DIFFERENT REINFORCEMENT CONDITIONS
ON THE TEST PERFORMANCE OF MULTIHANDICAPPED
DEAF CHILDREN¹*

PAUL M. SMEETS AND SEBASTIAN STRIEFEL

UNIVERSITY OF LEIDEN AND

BUREAU OF CHILD RESEARCH, UNIVERSITY OF KANSAS

The present study explored the effects of different reinforcement conditions on the number of correct responses on the Raven Progressive Matrices. Four groups of 11- to 18-year-old multihandicapped deaf children matched on the basis of mean age and pretest scores were used. The groups were randomly assigned to any of four posttest conditions: end-of-session reinforcement, noncontingent reinforcement, delayed reinforcement, and immediate reinforcement. The mean posttest score of subjects tested under the immediate-reinforcement condition was significantly higher than that of any other group. No significant differences were observed between the mean posttest scores of the three other groups. The practical implications of using reinforcement procedures for testing purposes are discussed.

DESCRIPTORS: deaf, test performance, psychological testing, Raven Progressive Matrices, effects of reinforcement on testing

The most common tool for assessing a person's level of cognitive functioning is the standard intelligence test. It has been recognized that test performance, and hence performance validity, is not independent of existing motivational conditions (Terman and Merrill, 1916; Thorndike, 1924; Wechsler, 1958). Studies designed to identify motivational variables that may increase test performance have used a variety of subjects, tests, and incentives. Most investigators (Ayllon and Kelly, 1972; Bergan, McManis, and Melchert, 1971; Edlund, 1972; Hurlock, 1925; Zigler and Butterfield, 1968) have reported that when responses are reinforced or when subjects were prevented from making many errors, test performance was higher than previous performance (and/or the performance of control subjects) under standard conditions. In contrast, Benton (1936), Maller

and Zubin (1932), and Tiber and Kennedy (1964) observed no significant difference in performance between subjects tested under standard conditions and those tested under reinforcement conditions. In fact, Roth and McManis (1972) observed that when tested on the Block Design Subtest of the WAIS, retarded adults with organic brain damage were less accurate when every correct response was reinforced than those who received no reinforcement.

No study has indicated what reinforcement procedure constitutes the optimal motivational condition. Instead, most studies focused on the effects of particular types of reinforcers, such as praise (Bergen, *et al.*, 1971; Hurlock, 1925; Roth and McManis, 1972; Tiber and Kennedy, 1964), tokens (Ayllon and Kelly, 1972), verbal reproof (Hurlock, 1925; Tiber and Kennedy, 1964), and candy (Edlund, 1972; Tiber and Kennedy, 1964) on test performance. In some of the studies (Bergan *et al.*, 1971; Edlund, 1972; Roth and McManis, 1972), subjects received reinforcement immediately after every correct response. In others (Ayllon and Kelly, 1972; Hurlock, 1925; Tiber and Kennedy, 1964), subjects received reinforcement after every subtest or test was completed; in other investigations (Benton, 1936; Maller and Zubin, 1932) sub-

---

¹This study was partially supported by NICHD grant HD 00870 to the Kansas Center for Research in Mental Retardation and Human Development, University of Kansas. The authors express their appreciation to Carmen Sandoval for his assistance and to Dr. Todd Rogers for his statistical advice. Reprints may be obtained from Paul M. Smeets, Department of Psychology, University of Leiden, Rijsburgerweg 157, Leiden, The Netherlands.

jects were promised reinforcement if they did better. In each of the above studies, motivation was defined in terms of one particular type of reinforcement procedure. The test results obtained under such conditions were compared with those obtained under conditions in which no reinforcers were delivered or promised. In essence, the design of previous investigations was based on a comparative evaluation of test results under two conditions only (*i.e.*, reinforcement *versus* no reinforcement). Thus, the specific effect of contingent reinforcement, noncontingent reinforcement, and immediacy of reinforcement upon the number of correct test responses has yet to be demonstrated.

The present study sought to analyze further which type of reinforcement contingency constitutes the optimal motivational condition as evidenced by the test performance of multihandicapped deaf children. More specifically, the study endeavored to compare deaf children's scores on the Raven Progressive Matrices when tested under four conditions: (1) end-of-session reinforcement, (2) noncontingent reinforcement; (3) delayed reinforcement, and, (4) immediate contingent reinforcement.

METHOD

Subjects

The initial group of subjects consisted of 52 deaf and hard-of-hearing children ranging from 11 to 18 yr of age. All children had been excluded from regular educational programs for the deaf, and participated in a special program designed to remedy their academic and behavioral deficits.

Procedure

Test instrument and testing procedures. All subjects were given the book form of the Raven Progressive Matrices. This test was designed to measure "observation and clear thinking" (Raven, 1960, 1965) and is often used with

deaf persons as a quick index to substantiate other, more comprehensive intelligence tests (Vernon and Brown, 1964). The test is composed of six sets (A, Ab, B, C, D, E) of 12 items each. Each problem consists of a large, uncompleted pattern with six (sets A, Ab, B) or eight (sets C, D, E) smaller patterns underneath it. The subject is requested to point at the one small pattern that will complete the large pattern. Under standard conditions, the scale is an untimed capacity test. A person's total score on all sets, except on set Ab, is used to assess a percentile grade.

For the purpose of the present study, sets A, Ab, and B were presented in color form (Raven, 1965); sets C, D, and E were presented in standard (*i.e.*, black and white) form (Raven, 1960). The response time allowed for each trial was 20 sec. Time was recorded by the experimenter counting silently (*i.e.*, 1001, 1002, *etc.*). A stopwatch was not used because of the different effects it could have had on the performance of the subjects (*e.g.*, some might try to hurry while others might not). Counting did not seem to interfere with test administration. Only the first response on each trial was scored. A correct response was defined as pointing to the correct pattern within a 20-sec interval after the problem was presented. Other responses, or the absence of a pointing response during the 20-sec interval, was scored as an incorrect response. The test performance was expressed in the total number of positive responses on all sets, including set Ab. During the first four trials (A₁ to A₄) the experimenter would point to the six small patterns and then to the missing part of the large pattern. In case the subject made an incorrect response or failed to respond, the experimenter would show the manual sign for "no" and again point to the small and large patterns. The experimenter would repeat this procedure until the subject made a correct response. From the fifth trial (A₅) on, the experimenter requested that the subject turn to the next page (problem) regardless of whether he had given a correct, an incorrect, or no response. The test

was discontinued after responding incorrectly on four consecutive trials.

Test Setting

Subjects were tested by two experimenters in a special testing room. Furniture included a desk, three chairs, and two closets. Experimenter and subject were seated at opposite sides of the desk, containing two test booklets, a cardboard screen, and scoring forms. Scoring of the subject's responses was conducted behind the screen, so that the experimenter's written responses were not visible to the subject. Sessions were held during school hours and lasted from 5 to 25 min. In case of reliability checks, the second experimenter would sit at one end of the desk; the screen did not allow this experimenter to observe the scored responses of the experimenter conducting the test. Children were always retested by the same experimenter. Only one of the experimenters was aware of the purpose of the study.

Design

A pretest/posttest control-group design (Campbell and Stanley, 1966) was used.

Pretest. The test was administered to all available ($N = 52$) pupils. Ten subjects were tested by Experimenter 1 and 42 by Experimenter 2. The test scores recorded by the two experimenters covered the same range. No instructions were given. Possible visual cues were avoided by having the experimenter look at the test booklet in front of the subject. At the end of the test session, the experimenter guided the subject to one of the closets containing cups with pennies and candy. All subjects were allowed to take 10

pennies, 10 small candies, or five big candies before leaving the room.

After the pretest was completed, subjects with a total score of five or less ($N = 7$) and 45 or more ($N = 1$) were excluded from further participation in the study. The remaining 44 subjects were then divided into four groups of 11 subjects each. Grouping was arranged so that the means ($\bar{X}$) and the standard deviations (SD) of the subjects' ages and pretest scores matched for all groups (Table 1). The four groups were then randomly assigned to any of four reinforcement conditions to be used during the posttesting.

Posttest. Retesting was initiated 17 days after pretesting began. With some exceptions, subjects were retested in the same sequence as during pretesting, so that the pretest/posttest interval was almost identical for all subjects. The posttest conditions for the groups were as follows: end-of-session reinforcement—Group 1; noncontingent reinforcement—Group 2; delayed reinforcement—Group 3; immediate reinforcement—Group 4.

End-of-session reinforcement. The procedures in this condition were identical to those used during the pretesting. The purpose of re-introducing this condition was twofold. First, it would demonstrate the effect of retesting on the test performance under identical conditions. Second, it would serve as a norm to compare the effects of the other conditions in which reinforcement procedures were used.

At the beginning of each session for subjects retested under the noncontingent-reinforcement, delayed-reinforcement, and immediate-reinforcement conditions, the experimenter showed the subject which items he could purchase for the

Table 1
Means ($\bar{X}$) and Standard Deviations (SD) for Age and Pretest Scores

	Groups							
	End-of-Session Reinforcement		Noncontingent Reinforcement		Delayed Reinforcement		Immediate Reinforcement	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Pretest Score	23.6	9.5	23.6	11.9	23.7	11.0	23.7	11.3
Age	14.5	1.6	14.5	1.5	14.4	1.9	13.8	1.9

checkmarks to be earned. A cardboard placed next to the subject indicated the value of each back-up reinforcer: three checkmarks for a penny or a small candy, six checkmarks for a big candy. The instructions were limited to the experimenter showing the manual sign for "three" while pointing to the small candy or penny, and the manual sign for "six" while pointing to the big candy. A special checkmark card was placed next to the subject. The experimenter counted the number of accumulated checkmarks at the end of the session. After the checkmarks were exchanged for candies or pennies, the subject would leave the room.

Noncontingent reinforcement. Subjects tested under this condition immediately received a checkmark for every (correct or incorrect) response, and a checkmark at the end of the 20-sec interval in case he failed to respond.

Delayed reinforcement. The procedures were identical to those in the end-of-session reinforcement condition except that, at the end of every sixth trial, the experimenter would request that the subject pause. The subject then received one checkmark for every correct response made during the preceding six trials, and then continued with the test.

Immediate reinforcement. Subjects tested under this condition received a checkmark immediately after every correct response.

Reliability. Agreement in scoring between the two experimenters was assessed during the pretest and posttest sessions. The number of reliability checks during pretesting and posttesting was 11 and 13, respectively. Agreement was based on item-by-item comparisons. Reliability was computed by dividing the number of observed agreements over the number of possible agreements, multiplied by 100; reliability during pretests varied between 94% and 100%, and between 98% and 100% during the posttests.

Data Analysis Procedures

To test the overall effect of the different reinforcement conditions, the posttest scores were

submitted to an analysis of covariance, using the age and pretest scores as covariates. The Tukey test was used for all analyses involving comparisons of group means.

RESULTS

The covariance analysis on the posttest scores revealed ($F = 6.64, 3/38 \text{ df}, p < 0.01$) that the mean scores of the four groups differed significantly. Descriptive data set forth in Table 2 show that the mean posttest scores of all four subjects exceeded the mean pretest scores. Little difference was observed between the mean pretest and posttest scores of three of the four groups. The per cent differences between the mean pretest and posttest scores were 2.1% for the end-of-session reinforcement group, 7.0% for the noncontingent-reinforcement group, and 8.8% for the delayed-reinforcement group (these percentages were computed by dividing the difference between the mean pretest and mean posttest score by the mean pretest score and multiplying the result by 100). In contrast, a pronounced difference (37.1%) between pretest and posttest performance was observed for the immediate-reinforcement group. The Tukey tests showed that the mean posttest scores of the three groups not receiving immediate reinforcement did not differ significantly from one another; the mean performance of the immediate-reinforcement group was significantly higher ($p < 0.05$) than that of any other group.

Despite the increased performances of all groups on the posttest, each group contained one or more subjects with a negative gainscore.

DISCUSSION

Present findings revealed that, of the four reinforcement conditions, immediate delivery of checkmarks for correct responses increased test performance most. This increase was largely contributed by five subjects whose gainscores exceeded those of any other subject in any other group. The total gainscore (88) of these five subjects accounted for 91.7% of the total gainscore

Table 2
Pre- and Posttest Scores of the Four Groups

S#	End-of-Session Reinforcement Group			Noncontingent Reinforcement Group			Delayed Reinforcement Group			Immediate Reinforcement Group		
	Pretest Score	Posttest Score	Gain	Pretest Score	Posttest Score	Gain	Pretest Score	Posttest Score	Gain	Pretest Score	Posttest Score	Gain
1	44	48	+ 4	43	46	+ 3	42	45	+ 3	44	50	+ 6
2	40	32	- 8	40	27	-13	40	38	- 2	38	38	0
3	30	24	- 6	37	38	+ 1	33	35	+ 2	34	37	+ 3
4	27	27	0	26	28	+ 2	27	29	+ 2	28	45	+17
5	26	32	+ 6	25	32	+ 7	26	27	+ 1	25	37	+12
6	22	24	+ 2	25	26	+ 1	24	27	+ 3	25	42	+17
7	21	19	- 2	19	27	+ 8	21	24	+ 3	21	13	- 8
8	17	24	+ 7	18	22	+ 4	18	26	+ 8	16	37	+21
9	16	15	- 1	13	12	- 1	15	15	0	15	20	+ 5
10	11	13	+ 2	8	14	+ 6	9	5	- 4	8	10	+ 2
11	6	7	+ 1	6	6	0	6	13	+ 7	7	28	+21
X	23.6	24.1	0.5	23.6	25.3	1.7	23.7	25.8	2.1	23.7	32.5	8.8
SD	9.5	10.6		11.9	11.0		11.0	11.0		11.3	12.5	

(96) of all 11 subjects tested under this condition. Of the remaining six subjects, fewer had a zero or negative gainscore than those of the other groups. It may be concluded that for these six subjects, the immediate delivery of checkmarks was not a controlling variable. The immediate delivery of checkmarks for correct responses probably also affected the rate of self-correction behaviors, even though self-corrections were not reinforced. Anecdotal observations indicated that subjects tested under this condition would frequently make a second, and often correct response, when they observed that they did not receive a checkmark for their initial response. Self-correction responses were rarely observed with subjects tested under the other conditions.

Present findings do not demonstrate whether the checkmarks served as conditioned reinforcers or merely as feedback. The subjects were not given direct experience in which checkmarks were linked with back-up reinforcers until it was too late to affect test performance. The subjects had been given nonverbal instructions that indicated that checkmarks could be exchanged for back-up reinforcers at the end of the session. This issue would have been clarified by linking the checkmarks with back-ups in an actual experience, or by including a group that received only checkmarks but with no back-up reinforcers given or mentioned. In either case, the checkmarks in the immediate reinforcement group did have a major effect on responding. It is possible that for five subjects in the immediate-reinforcement group, the checkmarks had both a reinforcing and feedback function, and for the other six only a feedback function, thus accounting for the difference in the performance of the subjects exposed to the same condition.

The delayed-reinforcement procedure was not expected to influence test performance as effectively as immediate reinforcement. However, in view of previous findings (Ayllon and Kelly, 1972; Hurlock, 1925) it was anticipated that delayed reinforcement would be more effective than noncontingent- and end-of-session reinforce-

ment conditions. It should be pointed out that in the Ayllon and Kelly (1972) and Hurlock (1925) studies, the effects of delayed reinforcement over no reinforcement could have been due to the *contrast* in reinforcement between these groups, since the no-reinforcement group did not receive reinforcement either before the posttest (history of reinforcement) or during the posttest (end-of-session reinforcement). In the present study, all four groups received reinforcement (pennies or candy) at the end of the pretest, thus potentially giving the subjects a history of reinforcement. Such a history of reinforcement, if it functionally existed, may have been sufficient to obscure the differential effects of the delayed-reinforcement condition, in contrast to the end-of-session and noncontingent-reinforcement conditions. In addition, it is possible that the contrast between contingent feedback (checkmarks) and a delay of even five trials was augmented by the intensity of the one-to-one contact between a child and the experimenter.

Subjects tested under the noncontingent-reinforcement condition received more checkmarks than those tested under the other conditions. Nevertheless, the test performance of these subjects was virtually identical to that of subjects who received no checkmarks until the end of the session.

Recent studies (Ayllon and Kelly, 1972; Edlund, 1972) underscored the importance of using reinforcement procedures to ensure that the child's test performance is an accurate and valid reflection of his abilities. Although test performance obtained using reinforcement procedures may reflect abilities, it remains unclear whether such results also reflect the child's typical level of classroom performance. Thus, it remains unclear whether test performances under reinforced or unreinforced conditions is most relevant when decisions involving long-term educational placement are based on the child's test score(s). Present findings suggest that the use of reinforcers, particularly when delivered immediately after every correct response, improves test performance. Although previous in-

investigators (Ayllon and Kelly, 1972; Hurlock, 1925) have demonstrated that delayed-reinforcement procedures may improve test performance, present findings suggest that the control exercised by this procedure may be weak and ineffective, compared to that of an immediate reinforcement procedure. Present observations also call for an explicit and precise description of reinforcement procedures when used in psycho-educational test situations. In many instances, progress in cognitive and academic skills is based on a pre-test/posttest procedure. Variations in the reinforcement procedures between pretest and posttest may result in invalid conclusions regarding the child's progress and/or the efficacy of the educational program.

REFERENCES

- Ayllon, T. and Kelly, K. Effects of reinforcement on standardized test performance. *Journal of Applied Behavior Analysis*, 1972, **5**, 477-484.
- Benton, A. L. Influence of incentives upon intelligence test scores of school children. *Journal of Genetic Psychology*, 1936, **49**, 494-497.
- Bergan, A., McManis, D. L., and Melchert, P. A. Effects of social and token reinforcement on WISC block design performance. *Perceptual Motor Skills*, 1971, **32**, 871-880.
- Campbell, D. T. and Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.
- Edlund, C. V. The effect on the behavior of children as reflected in the IQ scores, when reinforced after each response. *Journal of Applied Behavior Analysis*, 1972, **5**, 317-319.
- Hurlock, E. B. An evaluation of certain incentives used in schoolwork. *Journal of Educational Psychology*, 1925, **16**, 145-159.
- Maller, J. B. and Zubin, J. The effects of motivation upon intelligence test scores. *Journal of Genetic Psychology*, 1932, **41**, 136-151.
- Raven, J. C. *Guide to the standard progressive matrices*. London: Lewis, 1960.
- Raven, J. C. *Guide to using coloured progressive matrices*. London: Lewis, 1965.
- Roth, G. and McManis, D. L. Social reinforcement effects on block design performance of organic and nonorganic retarded adults. *American Journal of Mental Deficiency*, 1972, **77**, 181-189.
- Terman, L. M. and Merrill, M. A. *Measuring intelligence*. New York: Houghton Mifflin, 1916.
- Thorndike, E. L. Measurement of intelligence. *Psychological Review*, 1924, **31**, 219-252.
- Tiber, N. and Kennedy, W. A. The effects of incentives on the intelligence test performance of different social groups. *Journal of Consulting Psychology*, 1964, **28**, 187.
- Vernon, M. and Brown, D. A guide to psychological tests and testing procedures in the evaluation of deaf and hard-of-hearing children. *Journal of Speech and Hearing Disorders*, 1964, **29**, 414-423.
- Wechsler, D. *The measurement and appraisal of adult intelligence*. Baltimore: Williams and Wilkins, 1958.
- Zigler, E. and Butterfield, E. C. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. *Child Development*, 1968, **39**, 1-14.

Received 27 December 1973.

(Final acceptance 30 September 1974.)