DOCUMENT RESUME

ED 043 660 TM 000 081

AUTHOR Witmer, J. Melvin; Bornstein, Alan V.
TITLE The Effects of Verbal Approval and Disapproval Upon the Performance of Third and Fourth Grade Children on Four Subtests of the Wechsler Intelligence Scale for Children.
PUB DATE Mar 70
NOTE 18p.; Paper presented at the American Personnel and Guidance Association Convention, New Orleans, March 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$1.00
DESCRIPTORS *Elementary School Students, Grade 3, Grade 4, *Intelligence Tests, Interaction, Negative Reinforcement, Positive Reinforcement, *Rapport, *Reinforcement, Testing, *Verbal Operant Conditioning

ABSTRACT

Four selected subtests from the WISC, Arithmetic, Digit Span, Picture Arrangement and Block Design, were administered to 111 pupils in the third and fourth grades at Florida State University Elementary School. Social reinforcers, i.e., verbal approval and disapproval, were systematically presented to the examinees after specified items and between the subtests. Data was analyzed to determine different effects on test performance from the treatments and from the examiners. Results suggest that verbal behavior of the examiner can significantly influence the test performance of third and fourth grade boys and girls, particularly those reared in upper middle class families. Replication of this study is suggested t determine whether the results obtained hold true for children representative of the entire population and for children of different age levels. (PS)

# THE EFFECTS OF VERBAL APPROVAL AND DISAPPROVAL UPON THE PERFORMANCE OF THIRD AND FOURTH GRADE CHILDREN ON FOUR SUBTESTS OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

Paper Presented
at
The American Personnel and Guidance Association Convention
New Orleans, March, 1970

by
J. Melvin Witmer, Associate Professor of Education
Ohio University, Athens, Ohio

and

Alan V. Bornstein, Child Psychologist
Henry Pollak Memorial Clinic
Monmouth Medical Center
Long Branch, N. J.

# THE EFFECTS OF VERBAL APPROVAL AND DISAPPROVAL UPON THE PERFORMANCE OF THIRD AND FOURTH GRADE CHILDREN ON FOUR SUBTESTS OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

Since the advent of the individual intelligence test, the question of the effect of examiner-examinee interaction upon test performance has been raised. While recognizing that the test manuals provide stringent regulations to be followed in order to control and standardize the conditions for test administration, the variability of examiner-examinee interaction and its effect upon test performance requires more careful examination. The problem of the influence of different external factors on intelligence test results has been studied, but there are limited experimental data and conflicting results.

There is general agreement in the literature that the testing situation should elicit optimal performance of the examinee on a test of mental ability (Klugman, 1944; Terman, 1916; Terman and Merrill, 1937; Terman and Mer ill, 1960; Wechsler, 1949). Physical conditions and administration procedures can be specified and controlled, or at least noted when varying from normal. Although it is possible to describe the nature of the social relationship between the examiner and the examinee, it is difficult to control or assess the effect of this social relationship upon test performance in examiner-examinee interaction.

Terman as early as 1916 made the point that praise of a child's efforts in the testing situation contributed more than anything else to satisfactory rapport. The child should be kept interested, confident, and at his best level of effort. "Exclamations like 'Fine!,' 'Splendid!,' etc. should be used lavishly" during the examination (Terman, 1916,

p. 125). In a later publication, Terman and Merrill (1937) advised the examiner to enlist the subject's best efforts, otherwise the resulting score would be less than optimal to some unknown degree. The subject's best efforts are to be enlisted by the establishment and maintenance of adequate rapport. They felt that "it is wise to praise frequently and generously" (p. 57).

In their latest revision of the Stanford-Binet Intelligence Scale, Terman and Merrill (1960) reiterate the importance of rapport to elicit the subject's best efforts and maintain both high motivation and optimal performance level throughout the testing session. The examiner is advised to encourage the subject through frequent and generous praise, but this approval should be given for effort rather than success on a particular response. Spontaneous comments such as "Good!" and "Fine!" should be used to elicit the subject's best efforts. However, under no circumstances should the examiner show dissatisfaction with a response.

Wechsler (1949) in his Manual for the Wechsler Intelligence Scale for Children (WISC) is less clear and specific than Terman and Merrill (1960) in his consideration of the effect of examiner-examinee interaction upon the performance of the child in the testing situation. Like Terman and Merrill (1960) he hopes that the examiner will secure scores which represent the optimum ability of the child. In comparing his general testing considerations with those in the Stanford-Binet Manual (1960), it appears that he is encouraging an interpersonal relationship that is more neutral than that of Terman and Merrill. According

to Wechsler friendliness and warmth should characterize the examiner's approach; that in a school situation the examiner should be known and accepted by the child to be tested and by the other children in his group. In the testing situation supportive expressions are appropriate if the subject does not do well on a test. Yet, he cautions against the use of approval if a subject is making an effort or experiencing a modicum of success in the tasks.

> children vary in their reaction to commendation from an adult; in no case should the examiner indicate dissatisfaction with a response as given nor build up an expectancy of approval in the subject so that giving no comment would be interpreted by him as disapproval (Wechsler, 1949, p. 19).

Wechsler in his general testing procedures implies a somewhat neutral attitude as the proper frame of reference for the examiner in presenting the test materials. This neutral set may appear as a negative stimulus to some individuals. It is our contention that "no comment" type of examiner behavior in the testing interview is an ambiguous stimulus since children, as a result of prior social conditioning, come to the testing situation with the expectation of approval or disapproval from an adult. If an examiner does not use encouragement or express approval in some manner with some frequency, the examiner's "neutral" behavior is likely to be responded to in a variety of ways, thus the test results are more likely to be less than an optimal measure of intelligence.

Numerous studies have been conducted to observe the effect of material and social incentives upon human and animal behavior. Of particular interest to us were those studies which examined the effect of verbal incentives, i.e., praise and blame, upon the performance of school age children. Hurlock (1924; 1925a; 1925b) was one of the early psychol-

ogists to attempt to assess the effect of praise and reproof upon intelligence test performance in a testing situation. She first studied the effects of verbal incentives upon group intelligence test performance of third, fifth, and eighth grade children. In administering an intelligence test using the test-retest method with a one week interval, she used praise, reproof, and control groups. Her conclusion was that neither praise nor reproof was superior, but that both tended to result in better performance than did practice alone, with the treatment having no significant differences by age, sex, race, or intelligence. In a follow-up study (1925a) she again found that the verbal incentives of praise and blame tended to raise IQ scores more than practice.

Following Hurlock's early efforts, several researchers have tried to demonstrate the effects of incentives upon test performance. Klugman (1944), in using the 1937 edition of the Stanford-Binet Intelligence Scale, conducted a cross cultural study including Negro and white children. He found that, although not statistically significant, money was slightly more effective than verbal praise and that Negro children responded to the money incentive better than the white children. However, there was no control group; thus there was no means of determining to what extent incentives affected performance. Tiber (1963), in using the Stanford-Binet Intelligence Scale, Form L-M to study the effect of verbal incentives, found no evidence that they make a difference in test performance. The only statistically significant differences noted were those between the various class and caste groups which included middle and lower class white and Negro children. In a study by Willcutt and Kennedy (1963) with fourth grade children of lower-middle and upper-lower class,

praise was found to be more effective than either reproof or no incentive in performance on a discrimination task. There was no significant difference between level of intelligence and effectiveness of verbal incentives.

In a review of 33 praise and blame studies conducted over a fifty year period, Kennedy and Willcutt (1964) concluded that praise tends to facilitate learning and blame tends to have a debilitating effect. In what appears to be somewhat contradictory findings to the above studies, Marshall (1965) reviewed 32 incentive studies (only five of which were included in the Kennedy and Willcutt review) to assess the use of punishment incentives and reported that reproof is more effective in terms of its effects on subsequent performance when compared to praise and control, or neutral conditions. Perhaps this conflict can in part be understood on the basis of whether the reinforcement was scheduled directly following each student response or after the completion of the situation. As suggested by Cofer and Appley (1964), the effects of praise and reproof may differ according to whether the reinforcement is contingent upon reinforcing specific responses to a task or upon incentives that are provided to overall performance.

A review of the research within the verbal operant conditioning paradigm, in which experimenter variables were used as independent variables, provides evidence that the subject's verbal behavior can be manipulated by the experimenter's verbal behavior. The use of social approval through such generalized conditioned reinforcers as "Good" and "Fine" has been demonstrated to bring the verbal behavior of the subject under the control of the experimenter (Greenspoon, 1962; Kanfer, 1958; Krasner, 1958; Taffel, 1955; Verplanck; 1955; Hickes, 1956).

In these verbal conditioning studies, the reinforcement ranged from a minimal verbal cue such as "mmm-hm" to psychoanalytically derived interpretations. Verbal reinforcement by the experimenter was shown to affect the type of verbal behavior and the frequency of that verbal behavior such as an increase in self-reference statements, types of verbs used, sentence length, opinions stated, and use of personal pronouns.

Considering that the research indicates that verbal incentives can improve performance on visual-motor tasks and that one's verbal behavior can be systematically modified by a verbal reinforcer, we formulated a problem of trying to modify an individual's performance on subtests of the WISC through the use of social reinforcers by the examiner. Littell (1960) in a review of a decade of research on the WISC, stated that,

> there appears to be strong reason to suspect that WISC scores are affected systematically by many variables other than intelligence, but little information about the exact nature of these variables and the relationships involved is available (p. 153).

Littell further stated that specific research with the WISC is needed to study (1) variables in the relationship between examiner and examinee, and (2) the influence of circumstances upon the examination.

## HYPOTHESES

The purpose of this study was to determine the effects of three modes of test administration upon the performance of third and fourth grade boys and girls on four subtests of the WISC. Social reinforcers, namely, verbal approval and disapproval, were systematically presented to the examinees after specified items and between subtests to determine whether there was any differential effect on test performance.

The following null hypotheses were tested:

1. There is no significant difference between the mean scores of the group receiving Approval and the group receiving Disapproval on selected subtests of the WISC.
2. There is no significant difference between the mean scores of the group receiving Approval and the group receiving Neutral treatment on selected subtests of the WISC.
3. There is no significant difference between the mean scores of the group receiving Neutral treatment and the group receiving Disapproval on selected subtests of the WISC.

## PROCEDURE

The subjects for the study were 90 third and fourth grade pupils who were randomly selected from a total population of 111 pupils in the third and fourth grades at Florida State University Elementary School. They were randomly assigned to six groups identified in pairs as Disapproval, Neutral, and Approval (D, N, A). Table 1 shows the mean IQ, SD, and range for each group. A majority of the subjects were members of upper middle class families residing in residential areas in and around a capital city of 60,000 persons. Occupational designations of most of the fathers were in the professional, technical, and managerial categories.

Four selected subtests from the WISC were administered to all subjects: Arithmetic, Digit Span, Picture Arrangement, and Block Design. It was hypothesized that these subtests would be most sensitive to treatment effects and could be scored objectively.

Three treatments, namely, verbal disapproval (D), verbal approval (A), and neutral (N), were used. Verbal approval was defined by the statements "Good!," "Fine!," "That was good," "That was fine." Such statements were made after the subject's response to the first item (whether right or wrong) in each subtest and between subtests.

Disapproval was defined by the statements, "I thought you could do better than that" (after the response to the first item, whether right or wrong, in each subtest) and "That wasn't so good" (between each of the subtests). In all cases the statement was made while looking at the child. For the Neutral group, there was no conscious or scheduled attempt to provide systematic verbal approval or disapproval. The statement, "Now let's try these," was given to all subjects in the three groups immediately before starting the next subtest. Standard test administration procedures as prescribed in the WISC Manual were followed with all three groups. The three treatments were alternated so that no single treatment was limited to any given day or time of the day.

TABLE 1

EQUIVALENCE OF GROUPS FOR INTELLIGENCE[a]

| Group[b] | Mean IQ | SD | IQ Range |
|---|---|---|---|
| Disapproval (N=30) | | | |
| Examiner 1 | 112.4 | 10.6 | 95-133 |
| Examiner 2 | 112.3 | 10.3 | 98-131 |
| Total | 112.4 | 10.2 | 95-133 |
| Neutral (N=30) | | | |
| Examiner 1 | 112.9 | 13.1 | 93-132 |
| Examiner 2 | 112.7 | 10.1 | 97-134 |
| Total | 112.8 | 11.7 | 93-134 |
| Approval (N=30) | | | |
| Examiner 1 | 112.5 | 9.6 | 97-133 |
| Examiner 2 | 112.6 | 10.4 | 89-126 |
| Total | 112.5 | 10.1 | 89-133 |

[a]The Primary Mental Abilities Test was administered by Examiner 2 six weeks prior to the experiment.

[b]Each of the six groups had an N of 15.

Both examiners used the same physical facilities, but at different times, and followed standard procedures before and after the actual testing situation. Each examiner escorted the child from the classroom to the testing room which had a table and two chairs. While walking from the subject's classroom to the testing room, the examiner attempted to establish rapport with the child. A positive interpersonal relationship established through chit-chat before beginning an experiment is more likely to make the experimenter's use of reinforcement cues more effective (Sapolsky, 1960; Solley and Long, 1958). Before having the children in the D group return to their room, the examiner said, "You fooled me, you did better than I thought you would, in fact you did very well." No disturbing carryover of anxiety generated by the experiment was reported by the teachers.

## RESULTS

For the analysis of the data, the standard scaled scores for the WISC were used. The four subtest scores were summed and mean standard scores for each group (D, N, A) were computed (See Table 2). A one-way analysis of variance was used to test the significance of the differences among the means of the three treatments. The obtained $F$ ratio (5.05) was significant at the .01 level of confidence. (Critical $F$, with 2 and 87 degrees of freedom, was 4.85.) Tukey's HSD test was used to make all pairwise comparisons among means (Kirk, 1968). The results are shown in Table 3.

A two-way analysis of variance was made to determine whether there were any significant differences between examiners according to treatment administered (See Table 4).

TABLE 2

MEAN STANDARD SCORES AND STANDARD DEVIATIONS ON FOUR SUBTESTS OF THE WISC ACCORDING TO TREATMENT (D, N, A)

| Group[a] | Mean Standard Score | SD | Standard Score Range |
|---|---|---|---|
| Disapproval | | | |
| Examiner 1 | 42.9 | 7.9 | 30-61 |
| Examiner 2 | 40.7 | 6.3 | 27-50 |
| Total | 41.8 | 7.2 | 27-61 |
| Neutral | | | |
| Examiner 1 | 45.3 | 9.0 | 29-60 |
| Examiner 2 | 43.1 | 6.0 | 34-56 |
| Total | 44.2 | 7.7 | 29-60 |
| Approval | | | |
| Examiner 1 | 47.6 | 7.3 | 34-59 |
| Examiner 2 | 47.6 | 3.5 | 40-53 |
| Total | 47.6 | 5.7 | 34-59 |
| Grand Total | 44.5 | 7.3 | 27-61 |

[a]Each treatment group had an N of 30 with Examiner 1 and Examiner 2 each having an N of 15 in each treatment group.

TABLE 3

MEANS, DIFFERENCES AMONG MEANS, AND SIGNIFICANCE OF DIFFERENCES AT .05 LEVEL OF CONFIDENCE FOR THE APPROVAL, NEUTRAL, AND DISAPPROVAL TREATMENTS

| Means | Differences Among Means | | |
|---|---|---|---|
| | $M_A$ | $M_N$ | $M_D$ |
| $M_A$ = 47.6 | - | 3.4 | 5.8* |
| $M_N$ = 44.2 | | - | 2.4 |
| $M_D$ = 41.8 | | | - |

*$p < .05$ where HSD = 4.39, using a two-tailed test.

TABLE 4

ANALYSIS OF VARIANCE FOR TREATMENT (D, N, A) BY EXAMINER (1, 2)[a]

| Source | Sum of Squares | df | Mean Squares | F | Required F for $p < .05$ |
|---|---|---|---|---|---|
| Examiner | 48.4 | 1 | 48.5 | < 1 | 4.0 |
| Treatment | 504.9 | 2 | 252.4 | 5.0 | 3.1 |
| Interaction | 24.2 | 2 | 12.1 | < 1 | 3.1 |
| Within | 4272.9 | 84 | 50.9 | | |
| Total | 4850.4 | 89 | | | |

[a]Both independent variables (examiner and treatment) were assumed to be fixed.

## DISCUSSION

Examination of the mean standard scores of each group (D, N, A) indicates a differential effect between the type of treatment or verbal reinforcement and performance on the four subtests of the WISC (See Table 2). The verbal disapproval caused a decrement in performance and the verbal approval an increase when compared to a group which was given no such verbal reinforcement. Null hypothesis 1 was rejected since the group receiving verbal approval (A) scored significantly higher than the group (D) which received verbal disapproval. Although in actual practice it is not likely that this amount of negative verbal behavior would be present in the examiner, the finding does demonstrate the effect of such verbal behavior upon the performance of third and fourth grade children on standardized test items.

Null hypothesis 2 was not rejected since the group receiving verbal approval (A) did not score significantly higher than the group (N) which received no scheduled verbal reinforcement. The directional trend was in favor of the group that received verbal approval. If one extrapolated the average difference of .85 IQ points per subtest, the difference in comparing the two treatments for the complete WISC would be eight IQ points. Although not statistically significant in this experiment, this finding appears to have significance for further research since the treatment given these two groups may approximate the actual behavior of examiners who administer the WISC.

Null hypothesis 3 was not rejected since there was no significant difference between the performance of the subjects receiving no scheduled

verbal approval or disapproval (N) and those receiving verbal disapproval (D). While the difference was not accepted as significant, the results were in the direction of the N group performing higher. The "no comment" type of behavior in which no statements relevant to the performance were made positively or negatively to the subjects immediately before or during the testing period, could have been perceived by the subjects as ambiguity. However, the general effect presumably was more of a negative outcome in test performance than a positive one.

Although the results do not show any significant differences between examiners (See Tables 2 and 4), the trend is unidirectional for the D and N groups with no difference in the A groups. Examiner 1 is 11 years older and larger in build. Besides being chance differences, examiner differences (age, height, personality, etc) or the fact that many of the subjects knew Examiner 2 as a group test administrator in the school prior to the experiment may account for this trend. Kanfer and Karas (1959) report that the attitudes and perceptions of the subject toward the experimenter that are based upon a pre-experimental relationship can be expected to manifest themselves in the subject's responsiveness to the experimenter's verbal reinforcement on a subsequent conditioning task.

Subject responses to the verbal disapproval were varied. Some subjects reacted by responding more quickly to the items presented by the examiner while with others the time interval increased. Others imitated the speech of the examiner by word accent and volume. Some subjects showed evidence of physical changes in behavior such as increase in breathing rate, sighing, unproductive movements, shakiness, (especially

with the hands), whispered responses, a raise in the pitch of the voice, and putting the thumb in the mouth.

The results of this experiment suggest that the verbal behavior of the examiner within a testing situation can significantly influence the test performance of third and fourth grade boys and girls, at least of those who are reared in upper middle class families. Giving verbal disapproval and providing verbal approval have significantly different effects, with the latter resulting in higher test performance. While the social reinforcement given in real testing situations may not be as frequent or intense as in this study, the amount of approval (support, praise, or encouragement) given by different examiners is likely to vary considerably, and thus needs to be recognized as an examiner-examinee variable that can influence test results.

This study suggests areas for further research in examiner-examinee interaction and the differential effects of examiner and examinee variables upon test performance. Replicated studies should be conducted to determine whether these results hold true for children representative of our population as well as children of different age levels. Children from a different social-cultural milieu, for example, the culturally disadvantaged, should be included. Those children who have been trained by their parents through negative reinforcement might conceivably respond differently to the treatment in the testing situation than those who have been trained primarily through positive reinforcement. Additional treatment groups could be included to provide treatment levels of increased verbal approval and of increased verbal disapproval to determine whether there is a linear relationship. Further research might include other

subtests of the WISC, examiners of both sexes, several examiners and video tape recorded test sessions for analysis of verbal and non-verbal behavior. Other examinee variables that might be studied for differences according to treatment are level of anxiety, achievement level, and sex.

## REFERENCES


Cofer, C. N., & Appley, M. H. Motivation: theory and research. New York: Wiley, 1964.

Greenspoon, J. Verbal conditioning and clinical psychology. In A. J. Bachrach (Ed.), Experimental foundations of clinical psychology. New York: Basic Books, 1962.

Hurlock, E. B. The value of praise and reproof as incentives for children. Archives of Psychology, 1924, No. 71.

Hurlock, E. B. The effects of incentives upon the constancy of the I.Q. Journal of Genetic Psychology, 1925, 32, 422-434. (a)

Hurlock, E. B. An evaluation of certain incentives used in school work. Journal of Educational Psychology, 1925, 16, 145-159. (b)

Kanfer, F. Verbal conditioning reinforcement: schedules and experimenter influence. Psychology Reports, 1958, 4, 443-452.

Kanfer, F. & Karas, S. Prior experimenter-subject interaction and verbal conditioning. Psychology Reports, 1959, 5, 345-353.

Kennedy, W. A. & Willcutt, H. C. Praise and blame as incentives. Psychological Bulletin, 1964, 62, 323-331.

Kirk, R. E. Experimental design: procedures for the behavioral sciences. Belmont, Calif.: Brooks-Cole Publishing, 1968.

Klugman, S. T. The effect of money incentives versus praise upon the reliability and obtained scores of the revised Stanford-Binet test. Journal of General Psychology, 1944, 30, 255-259.

Krasner, L. Studies of the conditioning of verbal behavior. Psychological Bulletin, 1958, 55, 148-171.

Littell, W. M. The Wechsler intelligence scale for children: review of a decade of research. Psychological Bulletin, 1960, 57, 132-156.

Marshall, H. H. The effect of punishment on children: A review of the literature and a suggested hypothesis. Journal of Genetic Psychology, 1965, 106, 23-33.

Sapolsky, A. Effect of interpersonal relationships on verbal conditioning. Journal of Abnormal and Social Psychology, 1960, 60, 241-246.

Solley, C., & Long, J. When is "uh-huh" reinforcing? Perceptual and Motor Skills, 1958, 8, 277.

Taffel, C. Anxiety and conditioning of verbal behavior. Journal of Abnormal and Social Psychology, 1955, 51, 496-501.

Terman, L. M. The measurement of intelligence. Boston: Houghton Mifflin, 1916.

Terman, L. M., & Merrill, M. A. Measuring intelligence. New York: Houghton Mifflin, 1937.

Terman, L. M., & Merrill, M. A. Stanford-Binet intelligence scale. Boston: Houghton Mifflin, 1960.

Tiber, N. The effects of incentives upon intelligence test performance. Unpublished doctoral dissertation, Florida State University, 1963.

Verplanck, W. The control of content of conversation: reinforcement of statements of opinion. Journal of Abnormal and Social Psychology, 1955, 51, 668-676.

Wechsler, D. Wechsler intelligence scale for children, manual. New York: Psychological Corp., 1949.

Wickes, T. Examiner influence in a testing situation. Journal of Consulting Psychology, 1956, 20, 23-26.

Willcutt, H. C., & Kennedy, W. A. Relation of intelligence to effectiveness of praise and reproof as reinforcers for fourth graders. Perceptual Motor Skills, 1963, 695-697.