Reinforcement Effects on the I.Q. Scores of Institutionalized Children and Adults with Developmental Disabilities

Jody Robin Lewis
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Educational Assessment, Evaluation, and Research Commons, Psychoanalysis and Psychotherapy Commons, and the Special Education and Teaching Commons

Recommended Citation
https://scholarworks.wmich.edu/masters_theses/988

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
REINFORCEMENT EFFECTS ON THE I.Q. SCORES OF INSTITUTIONALIZED CHILDREN AND ADULTS WITH DEVELOPMENTAL DISABILITIES

Jody Robin Lewis, M.A.
Western Michigan University, 1991

Eighteen institutionalized school-age and adult subjects with severe/profound developmental disabilities were evaluated for the effects of self-selected reinforcement for correct responding on their I.Q. scores when tested with the Stanford-Binet Intelligence Scale (Terman & Merrill, 1960) or the Leiter International Performance Scale (Arthur, 1952). Subjects were tested under both standard and reinforcement conditions. The results indicated that overall, subjects when tested under reinforcement conditions showed a significant increase in mean I.Q. scores when compared with the same subjects tested one month previously under standard conditions. Several features of the data reveal intriguing results with regard to subjects and related preexisting or predetermined variables, including age, psychotropic medication, and test type. Results and implications of the data for I.Q. testing of institutionalized severely and profoundly retarded individuals were discussed.
ACKNOWLEDGEMENTS

I wish to express sincere appreciation to my advisor and committee chairperson, Dr. Jack Michael, for his inspiration, direction, and support in completion of my course of study; and to my committee members, Dr. David Lyon and Dr. Paul Mountjoy, for their guidance and assistance.

I am deeply indebted to Dr. Markley Sutton for his essential advice, encouragement, and technical assistance. Also, special thanks to Dr. Allan Cooper for his interest, succor, and counsel. Appreciation is also extended to Dr. Bruce Hesse and Dr. David Nolley for their helpful suggestions and to all others who helped along the way.

My deepest gratitude is extended to my family for their love and support—most especially: to my father for his eternal confidence in me; to my mother for her encouragement and sage advice; to my brother, Michael, for his special support; and to my son, Jordan, whose unconditional love provided the final impetus to bring this study to completion.

Jody Robin Lewis
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Reinforcement effects on the I.Q. scores of institutionalized children and adults with developmental disabilities

Lewis, Jody Robin, M.A.

Western Michigan University, 1991
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>METHODS</td>
<td>12</td>
</tr>
<tr>
<td>Subjects</td>
<td>12</td>
</tr>
<tr>
<td>Materials: Instruments</td>
<td>13</td>
</tr>
<tr>
<td>Procedure</td>
<td>16</td>
</tr>
<tr>
<td>Standard Condition</td>
<td>17</td>
</tr>
<tr>
<td>Reinforcement Condition</td>
<td>18</td>
</tr>
<tr>
<td>RESULTS</td>
<td>21</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>30</td>
</tr>
<tr>
<td>Age Group</td>
<td>32</td>
</tr>
<tr>
<td>Psychotropic Medication</td>
<td>33</td>
</tr>
<tr>
<td>Test Type</td>
<td>34</td>
</tr>
<tr>
<td>APPENDIX</td>
<td></td>
</tr>
<tr>
<td>A. Research Protocol Approval</td>
<td>38</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>41</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Descriptive Data by Individual Subjects ..... 14
2. Type and Dosage of Medication by Subjects
   on Psychotropic Medications .............. 15
3. I.Q. Data by Individual Subjects ........... 23
LIST OF FIGURES

1. Comparison of Previous I.Q. Scores ........ 22
2. Mean I.Q. Change: Total Sample .......... 24
3. Mean I.Q. Change: Age Group ............ 26
5. Mean I.Q. Change: Test Instrument ...... 28
INTRODUCTION

Early studies investigating the role of incentives in obtaining optimum standardized test scores emphasized the importance of a subject working at his highest level (Klugman, 1944). Terman and Merrill (1960) stated that if the examiner "has failed to enlist the subject's best efforts, the only thing certain is that the resulting score will be too low in some unknown degree" (p. 46). Yet, common standardized test administration procedure prohibits the use of any word or action which might indicate that the subject has made a correct or incorrect response once the test has begun (Terman & Merrill, 1960). Praise and encouragement may only be given for effort. The rationale for this practice is not entirely clear. According to Terman and Merrill, "To praise only successful responses may influence effort in the succeeding tests" (p. 51). Exactly why this "influence" is undesirable is never clearly stated. This is especially noteworthy as Terman and Merrill also emphasize the importance of eliciting the subjects' best efforts and not encouraging inferior types of responses. However, what is clearly stated is that "unless these standard procedures are followed the tests lose their significance" (p. 47). These procedures are considered "controlled conditions" without which estab-
lished norms have no meaning and test results are difficult to interpret.

If one views test performance as a set of behaviors (Bijou, 1971), then reinforcement should affect test performance as it would any other behavior. Numerous studies have indicated the efficacy of reinforcement procedures in raising standardized test scores. These studies are discussed below.

Ayllon and Kelly (1972) suggested that contingent reinforcement functions to maintain or increase motivation and, therefore, enhances test performance. They concluded that unless test performance is maximized in a standardized test situation, "the resulting test score must not be assumed to be a representative sample" (p. 483) of performance. They significantly raised the test scores of moderately retarded subjects on the Metropolitan Readiness Test using token reinforcement for each correct answer.

Edlund (1972) surmised that precise reinforcement procedures, including carefully chosen consequences need to be used in the testing procedure, "if one is to produce an accurate summary of the individual's learning progress or I.Q., which may be used as a basis for administrative decision" (p. 319). He found significant differences between the I.Q. scores of low-middle class children administered the Stanford-Binet under standard testing conditions and those who were given M & M candy for each correct response.
Several other investigators were successful in using reinforcement to raise standardized I.Q. test performance in varied populations.

Miller (1969) restored a young girl's performance on the WISC to the "normal" range of intellectual functioning, after her score had slipped to the moderately retarded range, by reinforcing each correct response with tokens equal to a penny apiece. Husted, Wallin, and Wooden (1971) found significant differences in both the I.Q. scores and mental ages on the Cattell examination procedure for one group of subjects. Subjects tested with M & M's added as reinforcement attained significantly higher I.Q. scores and mental ages than subjects tested with the standard Cattell procedure.

Several studies by Clingman and Fowler (Clingman & Fowler, 1975, 1976; Fowler & Clingman, 1977) examined the effects of contingent candy reinforcement on the I.Q. levels. Only initially low performing subjects significantly increased their I.Q. scores when tested under reinforcement conditions. These studies lend credence to O'Connor and Weiss (1974), who theorized that different populations given standardized tests demonstrate differential motivational deficits in test-taking situations, and that perhaps the application of reinforcement conditions might increase the I.Q. scores for those populations.

Dickstein and Ayers (1973) found that college women
taking the Wechsler Adult Intelligence Scale (WAIS) scored significantly higher after receiving incentive instructions than a comparable control tested under standard conditions. Subjects in the incentive group were told that the five best scorers would each receive one dollar. Control group subjects were administered the test without the monetary incentive.

In a study conducted by Kratochwill and Brady (1976), the test scores of "normal" adult undergraduates were significantly higher on the WAIS when given specific feedback for their responses (e.g., "completely correct," "was partially correct," or "incorrect"). Nonspecific praise was not effective in changing I.Q. performance. It was impossible to determine whether specific feedback functioned to reduce a motivational deficit, or had a discriminative function. The event of telling the subject that his answer was "completely correct" functioned as a more effective type of reinforcement than nonspecific praise, and/or it functioned as a discriminative stimulus for correct responding. Also, telling a subject that his answer was "incorrect" may have had a punishing effect (i.e., decreased incorrect responding).

Holt and Hobbs (1979) demonstrated that a response cost condition might be equally effective as reinforcement in producing higher test scores. According to a study by Steel and Barling (1977), contingent praise was effective
in improving the I.Q. test performance of mentally retarded children and contingent praise had discriminative properties which might have, at least in part, accounted for its effectiveness. Smeets and Striefel (1975) sought to determine which type of reinforcement contingency constituted the optimal motivational condition as evidenced by the test performance of multihandicapped deaf children. Immediate contingent reinforcement was most effective when compared with delayed reinforcement, noncontingent reinforcement, and end of session reinforcement (control condition), in improving test performance on the Raven Progressive Matrices. Control subjects' mean gain scores did not approach significance on the posttests.

Several studies have attempted to determine which form of reinforcement would be effective in improving the standardized I.Q. test performance of different subject populations (Fowler & Clingman, 1977; Goh & Lund, 1977; Klugman, 1944; Sheeckart & Bass, 1976; Steel & Barling, 1977; Terrell, Taylor, & Terrell, 1978; and others). For example, Quay (1971) attempted to evaluate the effects of two types of reinforcement (candy and praise) on the Stanford-Binet I.Q scores of black preschool children while simultaneously examining effects of two different types of communication on I.Q. scores. However, this study did not employ a control group, nor did it use a repeated measures design. So it was impossible to determine whether the
group similarities were present because the independent variables were equally effective or ineffective in optimizing the I.Q. scores of their sample.

When one considers the vast individual differences within populations with regard to what functions as reinforcement, it may prove more effective to select a type of reinforcement for each individual test subject rather than attempting to discover one form of reinforcement to use when testing an entire population.

Michael (1982) described an "establishing operation" as that which increases the effectiveness of some object or event as reinforcement or that which evokes behavior that in the past has been followed by the object or event. Thus, it is important to select a type of reinforcement based on each individual test subject's establishing operation with regard to the type of reinforcement. When using "conditioned reinforcement" such as praise, one would be advised to consider the individual test subject's history with regard to the object or event to be used as reinforcement.

Not all studies on the effects of incentive conditions or reinforcement on I.Q. scores have produced positive results. Benton (1936) tested normal third grade children on a group I.Q. test (Otis Self-Administering Test) under standard and incentive conditions. Children in the incentive group were given feedback on their next test scores
and promised a prize for improving their next test scores. No significant differences were present when the incentive group and control group were compared. In a more recent study, Clingman and Fowler (1975) investigated the effects of candy reinforcement on the I.Q. test scores of above-average first and second grade children. No significant differences were found between candy reinforcement contingent on correct responses, candy given noncontingently during the test, and a standard administration procedure. Sheckart and Bass (1976) examined the effects of praise, nonvocal reinforcements (e.g., a nod) contingent on correct response, or standard test conditions on the WAIS scores of black undergraduate students. No significant differences were found between the three groups. Goh and Lund (1977) investigated the effects of contingent verbal reinforcement on the I.Q. score of nursery school children of low and middle social-economic status. No significant effects due to verbal reinforcement were found on the Peabody Picture Vocabulary scores of these children. Each of the above studies that did not demonstrate a positive effect on I.Q. scores due to reinforcement, lacked an operation of any kind to help increase the effectiveness of the type of reinforcement used. Such an operation is desirable when examining the effects of reinforcement.

Another group of studies questioned whether significant increases in I.Q. scores due to reinforcement effects
are really meaningful (Busch & Osborne, 1976; O’Connor & Weiss, 1974). O’Connor and Weiss were also concerned that increases in I.Q. scores due to reinforcement would simply shift the distribution of I.Q. scores, thereby not affecting the relative position of the individual. Thus, increases in I.Q scores would be meaningless. But Clingman and Fowler (1975, 1976) and Fowler and Clingman (1977) showed that increases in I.Q. scores due to reinforcement did not occur for medium and high I.Q. populations. Willis and Shibata (1978) examined the effects of feedback and reinforcement on the test scores of preschool children of lower socioeconomic level. About half the subjects in each group fell into the classification of dull normal or below on the pretest. Classification based on posttest scores produced similar distributions for the control and feedback only groups. But all of the reinforcement group subjects classed dull normal or below on the pretest moved up to the level of average or above on the posttest. These studies suggest that O’Conner and Weiss’ fears were not well founded, and that statistically significant increases in I.Q. scores can be quite meaningful.

Despite the results of the above studies, norm-referenced I.Q. tests administered using standard procedures continue to be used to influence or decide the fate of children, students, and mentally retarded or developmentally disabled individuals, etc. Even the legal cutoff
point for federal aid for developmentally disabled persons is determined by an I.Q. score.

The validity and importance of I.Q. scores remain controversial. Recently, an entire issue of the Journal of Learning Disabilities was devoted to the subject (October, 1989). Siegel (1989) suggested that I.Q. test scores are not necessary for the definition of learning disabilities. Most I.Q. tests require specific knowledge and vocabulary that are usually acquired through school or reading. Persons without that history would likely score lower on their I.Q. tests, but are not necessarily less "intelligent." It is possible that because one of the major underlying assumptions of the Stanford-Binet and other similar tests is that the individual being tested has a "normal experiential background" (Budoff & Hamilton, 1976), experiential deficits may be responsible, in part, for low scores. Without the knowledge and skills common to persons of their own age, and with the prevalence of speech and language problems in the severe/profound population, the number of individuals in whom we can establish a "basal age" with the Stanford-Binet is limited. According to Budoff and Hamilton (1976), Stanford-Binet I.Q. scores were the least predictive of learning potential measures and of teachers' and staff's ratings of ability of three tests studied, yet they are still frequently used for this population.

Bryan (1989) suggested that current usage of I.Q. scores...
scores may interfere with the fundamental goal of helping children with learning disabilities. Torgensen (1989) was not sure that present knowledge justifies its use in the selection of children for special services. Graham and Harris (1989) stated that it is most important that we abandon I.Q. scores as decision makers regarding learning disabilities.

Lesak (1988) indicated that most psychologists, psychiatrists, educators, judges, the United States Social Security Administration, etc., think, write, talk and make decisions as if an I.Q. score represents real biological capacity, located inside the cranium, rather than considering the fact that I.Q. isn't a real entity. This fact has been typically and conveniently forgotten in the discussion of disabilities. Sheridan (1971) stated that all we really know is that responses on the tests are related to achievement on certain socially important tasks, but it is difficult to free ourselves from the conviction that academic performance is caused by a high I.Q. According to Zucker and Polloway (1987) current practice in the assessment of mental retardation and developmental disabilities includes two dimensions: general intellectual function as determined by I.Q. score, and adaptive behavior. Based on I.Q. score, level of retardation is assigned to the individual. Traditionally, and in many states and school systems today, categorization based on I.Q. scores decides the appropriate
provision of services.

The purpose of this study is to examine the effects of reinforcement on the I.Q. test performance of the lowest scoring testable group of institutionalized persons with developmental disabilities, using two common norm-referenced I.Q. tests. The role of preexisting and predetermined variables (i.e., age category, use of prescription psychotropic medication, and type of I.Q. test assigned) will also be examined.
METHODS

Subjects

Twenty subjects were selected from the population of severely and profoundly retarded residents of Oakdale Regional Center for Developmental Disabilities, Lapeer, Michigan.

The Oakdale Regional Center research committee reviewed and approved the participation of all subjects in the study (see Appendix A).

The Center's psychologists chose the subjects. The criteria for subject selection were that subjects must be: (a) testable with either the Stanford-Binet Intelligence Scale (Terman & Merrill, 1960) or the Leiter International Performance Scale (Arthur, 1952); and (b) had previously scored within the severe/profound I.Q. category on a standard measure.

In order for an individual to be considered for the study, he must have had in his record a previously recorded I.Q. score which indicated that he had passed the minimum year level (the II year level) on the Leiter or the Stanford-Binet and was adequately compliant with the procedures. Behaviors which would result in an individual being eliminated from the study were (a) not remaining on task
long enough for an examiner to complete at least one sub-
test, (b) attacking the examiner or throwing test materials
whenever testing was attempted, (c) or not emitting any
response to test questions or material. Subjects with
blindness, or physical handicaps, which precluded speech or
manipulation of test materials were also eliminated from
the study. The entire pool of eligible residents totaled
twenty-seven (27).

Of the 20 subjects selected, 2 were eliminated due to
the above stated conditions. Eighteen subjects completed
the study, 4 female and 14 male (see Table 1). Table 1
outlines the demographic information for the subjects.
There were 9 school-age subjects (under 26 years of age;
range 19-25) and 9 adult subjects (range 30-71 years).
Nine subjects were receiving psychotropic medication (see
Table 2) and 9 subjects were not. All subjects had been
institutionalized for a number of years (range 3-47 years;
mean = 18 years). All subjects selected had previous I.Q.
test scores reported as a part of the selection criteria.
These I.Q. scores ranged from 15-34; mean I.Q. score = 22.

Materials: Instruments

The Stanford-Binet Intelligence Scale (Terman &
Merrill, 1960) was selected because it is one of the most
frequently used individual norm-referenced I.Q. tests and
it is widely used with institutionalized developmentally
Table 1
Descriptive Data by Individual Subjects

<table>
<thead>
<tr>
<th>Subjct Number</th>
<th>Sex M/F</th>
<th>Age</th>
<th>Schl Age (≤25)</th>
<th>Years in Prev. Institun</th>
<th>I.Q.</th>
<th>Med. Y/N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>24</td>
<td>S/A</td>
<td>15</td>
<td>34</td>
<td>YES</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>23</td>
<td>S/A</td>
<td>5</td>
<td>15</td>
<td>YES</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>25</td>
<td>S/A</td>
<td>7</td>
<td>19</td>
<td>NO</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>19</td>
<td>S/A</td>
<td>14</td>
<td>26</td>
<td>YES</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>22</td>
<td>S/A</td>
<td>21</td>
<td>20</td>
<td>NO</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>24</td>
<td>S/A</td>
<td>2</td>
<td>15</td>
<td>NO</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>21</td>
<td>S/A</td>
<td>6</td>
<td>22</td>
<td>YES</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>22</td>
<td>S/A</td>
<td>10</td>
<td>15</td>
<td>NO</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>31</td>
<td>A</td>
<td>10</td>
<td>34</td>
<td>YES</td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>33</td>
<td>A</td>
<td>25</td>
<td>32</td>
<td>YES</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>49</td>
<td>A</td>
<td>32</td>
<td>15</td>
<td>NO</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>53</td>
<td>A</td>
<td>37</td>
<td>24</td>
<td>NO</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
<td>35</td>
<td>A</td>
<td>7</td>
<td>21</td>
<td>NO</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>71</td>
<td>A</td>
<td>12</td>
<td>18</td>
<td>YES</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>53</td>
<td>A</td>
<td>47</td>
<td>17</td>
<td>YES</td>
</tr>
<tr>
<td>16</td>
<td>M</td>
<td>55</td>
<td>A</td>
<td>47</td>
<td>25</td>
<td>NO</td>
</tr>
<tr>
<td>17</td>
<td>M</td>
<td>30</td>
<td>A</td>
<td>25</td>
<td>14</td>
<td>NO</td>
</tr>
<tr>
<td>18</td>
<td>M</td>
<td>24</td>
<td>S/A</td>
<td>9</td>
<td>25</td>
<td>YES</td>
</tr>
</tbody>
</table>

M=14  X=34.1  9 school age  X=18.4  X=21.7  9 yes  F=4  9 adult  9 no

* Psychotropic medication = Thorazine, Mellaril, or Haldol (see Table 2)

disabled persons (Budoff & Hamilton, 1976). The Stanford-Binet is regarded as a highly reliable test with most of the reported reliability coefficients for the various ages and I.Q. levels being over .90 (Anastasi, 1968).

The Leiter International Performance Scale (Arthur, 1952) was selected because it, too, is well recognized as useful with severely/profoundly retarded persons. Black (1973) reported a .916 test-retest reliability for the
<table>
<thead>
<tr>
<th>Subject Number</th>
<th>Type of Medications</th>
<th>Dosage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mellaril</td>
<td>50 mg. q.i.d.</td>
</tr>
<tr>
<td>2</td>
<td>Thorazine</td>
<td>50 mg. b.i.d.</td>
</tr>
<tr>
<td>4</td>
<td>Haldol</td>
<td>1 mg. q.i.d.</td>
</tr>
<tr>
<td>7</td>
<td>Thorazine</td>
<td>100 mg. t.i.d.</td>
</tr>
<tr>
<td>9</td>
<td>Haldol</td>
<td>5 mg. d.</td>
</tr>
<tr>
<td>10</td>
<td>Mellaril</td>
<td>100 mg. b.i.d.</td>
</tr>
<tr>
<td>14</td>
<td>Mellaril</td>
<td>75 mg. q.i.d.</td>
</tr>
<tr>
<td>15</td>
<td>Haldol</td>
<td>6 mg. d.</td>
</tr>
<tr>
<td>18</td>
<td>Mellaril</td>
<td>75 mg. q.i.d.</td>
</tr>
</tbody>
</table>

q.i.d. = 4 times/day  
b.i.d. = 2 times/day  
t.i.d. = 3 times/day  
d. = 1 time/day

Leiter over a 6-month period with 100 aphasic children. He also reported a mean test score increase from 81.59 to 83.44, a difference which did not approach statistical significance. Weiner (1971) recommended that the Leiter be regarded as satisfactory in reliability for group study. Bernsberg and Sloan (1951) found no significant differences between the Stanford-Binet and the Leiter when testing 55 brain-injured and 55 mental defective children. The Leiter
does not require the subject to be vocal, as the test is administered without words, nor does it require vocal expressions. Rather than relying on previously learned skills and knowledge to demonstrate I.Q., as is the case with the Stanford-Binet, the Leiter procedure requires demonstration of the correct response until the subject emits the correct response. Successive trials or subtests involve rapidly decreasing demonstration by the examiner while complexity gradually increases. These characteristics of the Leiter procedure make it especially useful with institutionalized persons with severe/profound developmental disabilities.

Procedure

This study was designed to maximize ecological validity by testing subjects in their own residential buildings, in familiar surroundings as their usual assessment would be conducted. All subjects received two administrations of a test. Both administrations for each subject used the same test and the same form. Five subjects were tested with the Stanford-Binet and thirteen subjects were tested with the Leiter. Subjects were assigned to the Stanford-Binet Test or the Leiter Test based on their psychologist’s recommendations. A one-month interval separated the first and second administrations of the tests.

The examiner was a master’s level psychology student.
trained in the proper administration of the Stanford-Binet and the Leiter. In order to measure examiner bias, inter-observer reliability checks were conducted during 14 of the tests (7 pretests; 7 posttests). Four of the Center's psychologists participated in the reliability check procedure. The same psychologist was present for each check during both the standard and the reinforcement conditions. The psychologist sat behind the examiner with a clear view of the subject. In this way, the psychologist was able to score the test independently of the examiner without prompting the scoring activity of the examiner.

**Standard Condition**

All subjects were tested under the standard condition first. Tests were conducted according to the procedures outlined in the testing manuals. Subjects were encouraged to participate with smiles and gestures for the Leiter and general words of encouragement during the Stanford-Binet ("Good going," "keep trying," etc.) according to the advice given in the test manuals. Encouragement was given while the subject was attending to the assigned task. Standard test procedure requires the examiner to estimate the appropriate level at which to begin the test for each individual. This is to reduce the time it takes to examine the subject and decrease the probability of boredom if the low level items are too easy. However, testing was begun at
the lowest level for all subjects and for both tests so as to insure that subjects not experience failure at a higher level at the beginning of the test. In view of the subjects' previous test scores, this procedure was appropriate and it also produced virtually identical testing conditions, except for experimental manipulation, for each subject. At the conclusion of each session, tests were scored in standard fashion according to the test manual instructions.

Reinforcement Condition

Types of Reinforcement

Several types of edibles were used as reinforcers. They included cheese puffs, circus peanut candies, caramel corn, sips of coffee, pieces of chewing tobacco, or M & M's.

Reinforcement Selection Procedure

In order to increase the probability that the edible items would be effective reinforcement, all testing took place three to four hours after the subject's previous meal. Subjects were tested at the same time of day for both standard and reinforcement conditions, and a tray was placed in front of each subject. Subjects were told, "We are going to be working on an activity now. Which of these
would you like to have some of while we work?" The sub-
jects then would reach for the edible item of his/her
choice and was allowed to consume it. This enabled the
examiner to watch to see that the subject consumed the cho-
sen item "eagerly" and served to "prime" the reinforcement.
The above procedures were conducted as establishing opera-
tions to assure the effectiveness of reinforcement. Whichever item the subject "chose" was used as reinforcement
throughout the I.Q. test. The chewing tobacco was an
exception to this procedure. It was recommended for one
nonvocal subject by the psychologist and had been previ-
ously approved for use as reinforcement in his habilitation
program.

All subjects were tested under the reinforcement con-
dition one month after being tested in the standard condi-
tion. Testing procedures were identical to the standard
condition, except as follows: Subjects were given the
instructions, "I am going to ask you to do some things for
me. If you do them correctly, you will get a piece of
this." The examiner pointed to the subjects' chosen edible
item. The testing was begun and subjects were then rein-
forced for each correct response. On the Stanford-Binet,
a "correct response" was defined as a correct answer or
combination of answers as defined in the test manual
instructions. For the Leiter, a "correct response" was
defined as the complete and correct configuration of blocks
as illustrated in the test manual. Subjects were allowed to consume the item immediately or to accumulate as they wished. At the conclusion of the test procedure, tests were scored in standard fashion according to the procedures in the test manual.
RESULTS

The subjects were selected, in part, based on previous I.Q. scores from various tests reported in the records (see Figure 1). The subjects were assigned to the Leiter or the Stanford-Binet tests by their psychologist. Subsequently, they were divided into Group L and Group S-B, i.e., the group receiving the Leiter and the group receiving the Stanford-Binet, respectively. There were no significant differences in mean I.Q. scores between the groups of subjects before the study, Group L mean I.Q. score = 22.46 and the Group S-B mean I.Q. score = 19.8 (Mann Whitney U=39; n1=5, n2=13, p>.10). The overall mean I.Q. score for the total sample for the previous I.Q. tests was 21.72.

The reliability checks for both the standard and reinforcement condition indicated little scoring bias. Scoring by the psychologists never deviated more than two I.Q. points from the examiner's scores. Of the fourteen testing sessions that were observed and scored (seven standard and seven reinforcement conditions), there was complete agreement in 10 of the sessions. Overall, the correlation between the examiner's scoring and the psychologist's reliability checks was .9.

Overall, there was a significant difference in I.Q. scores between the standard and the reinforcement.
Figure 1. Comparison of Previous I.Q. Scores.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
conditions. Table 3 displays the resulting I.Q. data by individual subject. In Figure 2, the change in mean I.Q. scores for the total sample is displayed. The change in mean I.Q. scores is from a mean I.Q. score of 20.77 to a mean I.Q. score of 28.55 (Wilcoxon T=9.5; N=18, p<.01).

The sample was divided into several different groups based on preexisting or predetermined factors, including age, the use of psychotropic medication and the type of

Table 3
I.Q. Data by Individual Subjects

<table>
<thead>
<tr>
<th>Subject Number</th>
<th>Type of Test</th>
<th>I.Q. Standard Condition</th>
<th>I.Q. Reinforce Condition</th>
<th>Difference (+ or -)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Leiter</td>
<td>36</td>
<td>47</td>
<td>11 +</td>
</tr>
<tr>
<td>2</td>
<td>Leiter</td>
<td>41</td>
<td>57</td>
<td>16 +</td>
</tr>
<tr>
<td>3</td>
<td>Leiter</td>
<td>24</td>
<td>22</td>
<td>2 -</td>
</tr>
<tr>
<td>4</td>
<td>Leiter</td>
<td>9</td>
<td>24</td>
<td>15 +</td>
</tr>
<tr>
<td>5</td>
<td>Leiter</td>
<td>36</td>
<td>43</td>
<td>7 +</td>
</tr>
<tr>
<td>6</td>
<td>Leiter</td>
<td>32</td>
<td>33</td>
<td>1 +</td>
</tr>
<tr>
<td>7</td>
<td>Leiter</td>
<td>22</td>
<td>32</td>
<td>10 +</td>
</tr>
<tr>
<td>8</td>
<td>Leiter</td>
<td>22</td>
<td>28</td>
<td>6 +</td>
</tr>
<tr>
<td>9</td>
<td>Leiter</td>
<td>0</td>
<td>37</td>
<td>37 +</td>
</tr>
<tr>
<td>10</td>
<td>Leiter</td>
<td>32</td>
<td>37</td>
<td>5 +</td>
</tr>
<tr>
<td>11</td>
<td>Leiter</td>
<td>0</td>
<td>7</td>
<td>7 +</td>
</tr>
<tr>
<td>12</td>
<td>Leiter</td>
<td>24</td>
<td>30</td>
<td>6 +</td>
</tr>
<tr>
<td>13</td>
<td>Leiter</td>
<td>15</td>
<td>27</td>
<td>12 +</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>S.B.</td>
<td>13</td>
<td>14</td>
<td>1 +</td>
</tr>
<tr>
<td>15</td>
<td>S.B.</td>
<td>13</td>
<td>20</td>
<td>7 +</td>
</tr>
<tr>
<td>16</td>
<td>S.B.</td>
<td>20</td>
<td>17</td>
<td>3 -</td>
</tr>
<tr>
<td>17</td>
<td>S.B.</td>
<td>14</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>S.B.</td>
<td>21</td>
<td>24</td>
<td>3 +</td>
</tr>
</tbody>
</table>

X=19.8          X=29          X=10.06 +

X=16.2          X=17.8        X=1.6 +

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 2. Mean I.Q. Change: Total Sample.
test administered. The results of this examination are reported below.

The sample was divided into school-age and adult subjects (see Figure 3). Figure 3 displays the results of the two test conditions and the two subject age groups. The school-age subjects' mean I.Q. scores showed a significant change from 27 to 34.44 (Wilcoxon T=2, N=9; \( p < .01 \)) from the standard to the reinforcement test condition. The adult subjects' mean I.Q. scores showed a similarly significant change from 14.55 to 22.55 (Wilcoxon T=3, N=9; \( p < .02 \)).

The subjects were also divided into two groups, based on their use of psychotropic medication. The results of this examination are displayed in Figure 4 (see Figure 4). The group of subjects that received psychotropic medication and the group of subjects not receiving psychotropic medication each had a mean I.Q. score of 20.77 for the standard test condition. The subjects receiving psychotropic medication had a significant increase in mean I.Q. score in the reinforcement condition to 32.44 (Wilcoxon T=0, N=9; \( p < .01 \)). The mean I.Q. for the group not receiving medication was 24.55 in the second condition (Wilcoxon T=7, N=9; \( p > .05 \)).

With regard to the use of different I.Q. tests, the results are displayed in Figure 5 (see Figure 5). The subjects assigned to the Leiter showed a significant change in mean I.Q. scores from 22.54 to 32.62 (Wilcoxon T=2, N=13,
Figure 3. Mean IQ Change: Age Group.
Figure 4. Mean I.Q. Change: Medication.
Figure 5. Mean I.Q. Change: Test Instrument.
The subjects completing the Stanford-Binet showed a nonsignificant change in mean I.Q. score from 16.2 to 17.8 (Wilcoxon T=3, N=5, p<.05).

Given the limitations required by the selection criteria, there is little reason to make assumptions regarding the normality of the distribution of the population examined and the type of measurement utilized. Furthermore, since the examination of the subject sample required several preexisting or predetermined variables to be utilized, it was decided that nonparametric statistics would be appropriate. It should be noted that the use of nonparametric statistics (Siegel, 1956) limits the degree to which the results can be generalized beyond the immediate sample.
DISCUSSION

The hypothesis that institutionalized subjects with severe or profound developmental disabilities would significantly improve their I.Q. scores when reinforced for correct responding was supported. A significant feature included in the reinforcement condition was the use of self-selected reinforcement. The results indicate that there was a significant increase in I.Q. scores when self-selected reinforcement was added to the standard test procedures of two norm-referenced I.Q. tests. Due to constraints placed on the study by the institution, no more than two trials of the I.Q. tests were allowed, thus, research designs which control well for the effects of history, such as ABA or Multiple Baseline designs, could not be used. However, it is still unlikely that this increase was due to history, as the Leiter has a test-retest reliability coefficient of .916 and the Stanford-Binet of at least .90.

There are several features of these results that might be evaluated. It should be noted that of more than 1000 individuals living at Oakdale Regional Center only 27 clients were identified as acceptable subjects. Twenty subjects were selected and, of these, 18 subjects completed the tests. It is clear that the results of this study are
circumscribed by the particular requirements of subject selection in this study. The study focused on those individuals who would be considered at a primary level for standardized I.Q. tests. With these findings, there is evidence that the effects of reinforcement on standardized I.Q. scores may be extended to individuals not previously considered. Furthermore, these effects may have specific value to clients previously considered to have scored in an I.Q. range "too low" for examination.

Further evaluation of the effects of reinforcement procedures appears to be indicated. The reinforcement selection procedure and testing schedule outlined in this study were effective in producing a type of reinforcement that would increase overall the I.Q. scores of the severely/profoundly developmentally disabled subjects. This supports the contention that it is not necessary to devote much time and research to the problem of determining what type of reinforcer works best for which population. Instead, a procedure which allows individuals to select their own type of reinforcer might be the most efficient way to address this problem.

Several features of the study displayed intriguing results with regard to the subjects, and related preexisting and predetermined variables.
Age Group

When the subjects were grouped and analyzed in two distinct categories (i.e., school age and adult), both groups' I.Q. scores improved significantly when self-selected reinforcement was added to the standard test procedure. Subjects in the school age category began with higher mean I.Q. scores in the standard test condition, and improved significantly. The adult group's mean I.Q. scores in the standard condition were lower, but also improved significantly, paralleling the results of the school age group. These data lend support to the hypothesis that both school age and adult subjects' I.Q. scores would improve significantly when self-selected reinforcement was added to the standard test procedure. The mean age of the school-age group was 22.67 and the mean age of the adult group was 45.55. Not only is this actual age difference a significant factor to consider, the effects of school programs for the young and changes in public policies over the past twenty years may be factors to be evaluated. It should be noted that the scores of the school-age clients and the adult clients each improved by the same degree. This indicates that reinforcement had effects on the responses of each group. Since the mean scores for the standard and reinforcement conditions for the adult group are each lower than the school-age group's mean I.Q. scores (i.e., the
lines are parallel), this could indicate a difference in groupings or a difference in skills. One is reminded, however, that the selection criteria applied to all the subjects. Given the change in resources to the population of persons with developmental disabilities over the past twenty years, the overall suppression in scores for adults at the institution as compared to school-age clients would constitute an interesting extension of this research.

Psychotropic Medication

Subjects taking prescription psychotropic medication, predetermined by their physician, were analyzed separately from the subjects not prescribed psychotropic medication. Improvement of their I.Q scores when tested in the reinforcement condition was explored. The nine subjects taking prescription psychotropic medication obtained significantly improved I.Q. scores when self-selected reinforcement for correct responses was added to the standard test procedures. The nine subjects not taking psychotropic medication did not improve their I.Q. scores significantly when tested in the reinforcement condition. The use of psychotropic medication was determined by the center's physicians and was not a variable that could be "manipulated" or assigned randomly. Since this effect is predetermined or "naturally occurring," it is difficult to determine, from these data, the exact effect of the use of psychotropic
medication. There are several possibilities. For example: (a) The subjects may derive from two distinct populations which are differentially affected by self-selected reinforcement, or (b) the psychotropic medication may somehow serve as an establishing operation by increasing the sensitivity to or the effect of the self-selected reinforcer in increasing I.Q. scores for this sample population, or (c) the psychotropic medication may have enabled the users to focus more readily on the environmental events and consequences regarding their responses.

Test Type

The hypothesis that I.Q. test scores would improve significantly for subjects taking either I.Q. test, when self-selected reinforcement was added to the standard testing procedure (reinforcement condition), was not supported by these data. The Group L subjects did improve significantly and the Group S-B subjects did not improve during the reinforcement condition. With the absence of a control group, it may be argued that significant improvements of Group L subjects were merely due to the second administration of the test. However, previous studies indicate that second administrations of the Leiter (Black, 1973) and other standardized tests (Ayllon & Kelly, 1972; Smeets & Streiffel, 1975) do not by themselves result in significant increases in test scores.
One explanation for the absence of a significant increase in I.Q. scores for Group S-B subjects under reinforcement conditions could be that reinforcement may not have occurred with sufficient frequency to improve performance. On the whole Group S-B subjects made more errors and received fewer reinforcers than did Group L subjects. Another factor may also have contributed to this outcome. Subjects in the reinforcement condition had already selected and sampled the edible items before the beginning of the test. These items remained in sight throughout the test. Errors and the subsequent withholding of reinforcement could have had a punishing or distracting effect on their performance. This could have competed with the reinforcing effect of the edible items. It is less likely that this would occur at the beginning of the Leiter test. This is because subjects tested with Leiter are merely expected to imitate the examiner as their first correct responses.

It is not entirely clear why severely/profoundly developmentally disabled subjects who were examined with the Stanford-Binet did not significantly improve their I.Q. scores when tested under reinforcement conditions. This apparent discrepancy warrants further investigation.

Inherent in all clinical or applied research are ethical or situational conditions which may compromise the integrity of the research design. Design considerations and decisions must be made within the constraints of the
applied research setting. The present study was conducted at a large state residential facility for persons with developmental disabilities undergoing habilitation programs. Their regular assessment process included I.Q. testing. Permission to conduct the study was given by the center’s research committee only if the subjects were identified by the center’s psychologists, and that no more than two trials of each I.Q. test be given per subject so as to interfere minimally with the subjects’ regular assessment. Only 27 subjects were identified by the psychologists for inclusion in the study (meeting the criteria) and 20 were selected to be tested. Eighteen completed both tests. Given the small sample size, nonparametric statistics were utilized. With these statistics, generalization was somewhat limited; however, the assessment of significance was possible.

This study has outlined several significant effects of reinforcement on the I.Q. scores of a sample of persons with developmental disabilities. Within the sample of individuals, numerous specific, interesting effects of the use of reinforcement on I.Q. test performance have been demonstrated. It is clear that as long as norm-referenced I.Q. tests continue to be administered for the purpose of diagnosis, profiling, and placement of individuals, there is a need to support each individual’s best effort and continue to scrutinize the factors which affect each person’s
outcome. The results of this study suggest that it is important that an I.Q. test be chosen which is appropriate, i.e., makes the correct assumptions about the basic skill levels of the subjects being tested, and that procedures that are used should be evaluated to ensure that the resulting I.Q. scores reflect as closely as possible the individuals' optimum performance. Further investigation of the issues outlined may lead to valuable uses of I.Q. scores and test procedures for individuals that have heretofore been excluded from consideration. The use of I.Q. scores as determinants in the allocation of resources, placement of individuals, and the implementation of treatment plans, becomes circumspect as research outlines the factors that may significantly affect the outcome.
Appendix A

Research Protocol Approval
MEMORANDUM
June 1, 1978

TO: Jody Lewis, Staff Development
FROM: David A. Ethridge, Ph.D.
Facility Director
SUBJECT: Research Proposal

Your research proposal entitled "The Effects of Reinforcement Procedures on the IQ Scores of Institutionalized Severe and Profound Retardates", as recommended by the Research Committee, has been approved.

DAE:mkk

cc: Dr. Berker
June 20, 1991

To Graduate College, Western Michigan University

From Jack Michael, Professor, Psychology Department

About Use of human subjects in the thesis research of Jody Lewis

Ms. Lewis did the research reported in this thesis in 1981 when the role of our own Human Subjects Institutional Review Board had not been established. I was her thesis advisor during that period and I was very familiar with the nature of her research activities. I also knew that she had obtained approval from the human subjects committee at the facility where the research was taking place. I am quite confident that there were no risks to the subjects from this research activity. From their perspective they were simply participating in another of the many training activities or assessment activities that occur at such a facility.
BIBLIOGRAPHY


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Terrell, J., Taylor, J., & Terrell, S. (1978). Effects of type of social reinforcement on the intelligence test


