Order Number 9020796

Incentive conditions and the selected WISC-R subtest performance of elementary school children

Pollock, Mary-Georgia Ann, Ph.D.

City University of New York, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

INCENTIVE CONDITIONS AND THE SELECTED WISC-R SUBTEST PERFORMANCE OF ELEMENTARY SCHOOL CHILDREN

by

Mary-Georgia Ann Pollock

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1989

Approval Page

This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

December 18, 1989
Date

Philip A. Saigh
Chair of Examining Committee

December 18, 1989
Date

Carol Kehr Tittle
Executive Officer

Alan L. Gross

Barry J. Zimmerman

David Goh

Stanley Jacobs

Supervisory Committee

The City University of New York

Abstract

INCENTIVE CONDITIONS AND THE SELECTED WISC-R SUBTEST PERFORMANCE OF ELEMENTARY SCHOOL CHILDREN

by

Mary-Georgia Ann Pollock

Advisor: Professor Philip A. Saigh

There is a general accord in the literature of individual intelligence testing regarding the need to facilitate the examinee's performance. Among the techniques employed to attain this goal have been operant procedures (e.g., verbal praise for effort or performance) and symbolic modeling techniques (e.g., filmed or videotaped vignettes of models' interactions with examiners). The present study compared the effects of two types of examiner comments (i.e., non-contingent verbal praise and neutral, non-evaluative comments) and three types of modeling experiences (i.e., videotaped modeling of test behavior, accompanied by onscreen narration; videotaped modeling of test behavior without narration; and no modeling), on the WISC-R Arithmetic, Picture Completion, Block Design, and Digit Span subtest scores of fourth graders. The 96 primarily white (97%) middle class students attending urban Roman Catholic

parochial schools, were matched by sex and scores on the Otis-Lennon School Ability Test (OLSAT), Form R, to one of the six experimental treatments. A multivariate analysis of covariance (MANCOVA) was applied to the test scores. The OLSAT School Ability Index scores served as the covariate. Three hypotheses were tested. H_{01} predicted a main effect for examiner praise, i.e., that the mean scaled scores of subjects who were verbally praised would significantly exceed the mean scaled scores of subjects who were not praised. H_{02} predicted that the mean scaled scores of subjects who viewed modeling videotapes would significantly exceed the mean scaled scores of subjects who did not view a modeling videotape. H_{03} predicted that the mean scaled scores of subjects who viewed the modeling videotape with narration would significantly exceed the mean scaled scores of subjects who viewed the modeling videotape without narration. In addition, an interaction effect was predicted. Data analysis failed to support any of the predictions. Several reasons were discussed regarding the non-significant results. Specifically, the verbal praise statements might have been poor motivators. In addition, subjects' attributions for success, sample characteristics, and duration of treatments might have affected the outcome measures.

ACKNOWLEDGEMENTS

In 1964, I attended a health careers conference for high school students at Kings County Hospital. Impressed by a presentation on the role of the child psychologist, I decided to pursue that career. During the intervening years, many individuals have helped me in my quest for that goal. Initial gratitude is due my parents, Elizabeth Vesey and Joseph Paul Pollock. At a time when society generally ignored girls' academic achievement and career aspirations, my parents were steadfast in telling me that "You could do anything if you set your mind to it." They were truly non-sexist or gender-blind long before those expressions entered the vernacular.

I am also grateful to Professors Sybil Gottlieb, William Greenstadt, and Robert Schwartz of the School Psychology Program at the City College of New York. They not only honed my clinical skills, but actively supported my pursuit of the doctorate. I am also deeply indebted to Professor Philip A. Saigh, chair of my dissertation committee, whose guidance and indefatigable efforts enabled me to complete the dissertation in a timely fashion. In addition, I am grateful that Professors Alan Gross and Barry J. Zimmerman served as members of my dissertation committee. Their expertise

elucidated critical issues in methodology and theory. I also wish to thank Professor David Goh and Dr. Stanley Jacobs who served as readers. Their comments were most welcomed.

Many doctoral candidates would agree that finding subjects for a study can be a formidable undertaking. That task proved relatively easy for me, thanks to the cooperation of the following principals and grade coordinators in the Roman Catholic Diocese of Brooklyn: Sisters Mary Ellen Malloy, Maureen McDermott, and Mary Helen Slavinskas; Ms. JoAnn DiGangi, Mr. Dennis Farrell, and Ms. Marjorie Novak.

I am also grateful to friends and classmates who rejoiced in my intermediate successes and lent sympathetic ears when my spirits were low. In particular, I would like to thank Jeanne McGuirk Beltran, Ph.D., whose own trek down the same academic path as mine served as a vicarious modeling experience. Finally, I would like to thank Robert Redfield whose humor and thoughtfulness were greatly appreciated.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

There is a general accord in the literature of intelligence testing regarding the need to obtain an examinee's best effort during standardized testing situations (Anastasi, 1987; Binet & Simon, 1983; Cronbach, 1984; French & Murphy, 1985; Sattler, 1988; Terman & Merrill, 1973). Cronbach (1984), for example, observed that

> the distinguishing feature of a test of ability is that the test taker is encouraged to earn the best scores he can. The goal of the test should be to bring out the person's best possible performance (within the rules); this means that the examinee must want to do well and must understand what is considered a good performance (p.29).

Maximal performance is not an end in itself but a means to increase the likelihood that valid test interpretations are made from such scores. Aware of the importance of test interpretation, Reschly (1979) argued that "If the child cannot or does not perform as well as possible owing to unique features of the testing environment, the results of the test are inaccurate reflections of the child's thinking competencies or academic skills (p.232)." Thus, securing scores that are indicative of an examinee's maximum effort has relevance to the practice of school psychology inasmuch as test scores represent a major factor in decisions

regarding diagnosis, intervention, and class placement (Weinberg, 1989).

Selected Review of the Literature

Viewed from an applied perspective, psychologists have generally sought to facilitate an examinee's efforts during standardized testing situations through the use of operant procedures (e.g., verbal praise contingent on effort or performance) and modeling techniques (e.g., filmed or videotaped vignettes of children's interactions with examiners). This selected review of the literature will provide a broad overview of the studies that have used either reinforcement or modeling paradigms to facilitate achievement on standardized measures of mental ability.

Verbal Praise

The assumption that verbal praise facilitates examinee performance is well entrenched in the annals of cognitive assessment. Binet and Simon recognized the value of praise as early as 1916 and offered the following advice for the assessment of mentally retarded students:

> For anyone who knows these poor beings and to what degree they open out when praised and on the contrary how quickly they withdraw within themselves at the least reproof, there is no doubt that this indulgent attitude is indispensable to obtain on their part even a small output of effort (p.140).

In a similar vein, Terman (1916) indicated that "nothing contributes more to a satisfactory rapport than praise of a child's efforts.... Statements like 'Fine!' 'Splendid!' etc. should be used lavishly (p.125)." Analogously, Wechsler (1974) advocated "judicious reinforcement" (p.55) of an examinee's efforts throughout the course of testing. One must keep in mind, however, that Wechsler did not provide an operational definition of "judicious reinforcement." More recently, Kaufman and Kaufman (1983a) instructed examiners to "maintain a high level of motivation by giving appropriate praise and encouragement.... Praise children for their effort, not the correctness of their responses" (pp. 13-14). They recommended such interjections as "Nice try," "Good," "You're working hard," or "That was a tough one." On the other hand, they cautioned against overworked expressions following each item because such comments serve no useful function.

The use of praise as a vehicle for enhancing performance on cognitive tests has been the subject of considerable study. Viewed in this context, Kennedy and Willcutt (1964) analysed the literature on praise from 1914 through 1964. Their review indicated that praise generally facilitated performance. On the other hand, a

more recent analysis by Fish (1988) failed to denote a clear relation between verbal praise and enhanced performance on various measures of mental ability. The lack of reviewer concordance is understandable for two reasons. First, Fish's critique spanned a greater period of time than that of Kennedy's and Willcut's analysis. Second, the reviews differed in scope. More specifically, Kennedy and Willcut examined the consequences of praise during individually and group administered tests. Moreover, Kennedy's and Willcut's analysis included cognitive and academic achievement measures. Fish, on the other hand, appraised research on "rewards" (which included not only praise but candy, tokens, toys, and knowledge of test results). In addition, Fish confined his review to studies of non-special education children during individual intelligence testing sessions.

Insofar as these reviewers took different tacts, a more circumscribed analysis of the effects of verbal praise during the course of individually administered intelligence tests with school aged children may serve to clarify the discrepancies between the previous reviews. In so doing, computer and manual searches were conducted. Two databases were used: the Educational Resources Information Center (ERIC) and PsychINFO. Both

searches spanned the literature from 1974 (the year of the publication of the Wechsler Intelligence Scale for Children - Revised) through October 1988. The library search consisted of examination of Psychological Abstracts for the same time period. References appearing in Kennedy and Willcutt, and in Fish were checked. In addition, bibliographies of the various Wechsler scales, the fourth edition of the Stanford-Binet, and the Kaufman Assessment Battery for Children that were published in The Ninth Mental Measurements Yearbook (Mitchell, 1985) were surveyed. This process yielded 24 pertinent studies.

Historical Contributions

Strictly speaking, the contributions of Hurlock (1924, 1925) and Benton (1936) are outside the realm of this review because they employed group intelligence measures. These investigations, however, merit recognition because they introduced praise as a manipulated variable during the intellectual assessment of school children. Previously, the outcome variables were scores attained on achievement tests.

In her initial study, Hurlock (1924) matched 408 children on the basis of sex, age, race, and scores on the National Group Intelligence Test, Scale B, Forms 1 and 2. (The author and year of publication were not

stated.) Subjects were subsequently randomly assigned to one of three treatment conditions (i.e., praise, reproof, or no feedback). Gain scores comprised the dependent variable. Hurlock observed that both praise and reproof were superior to the control condition. A year later she replicated the study with 132 fifth and 141 eighth grade students (number of boys and girls unstated) and similar results were observed.

Verbal Praise During Individual Testing

In Benton's classic study, incentives (a combination of praise, promise of a prize, exhortation, and knowledge of results) and standard test administration were compared. Twelve male and 13 female junior high school students completed the Otis Self Administering Test. (The publication date was not reported.) The school principal presented the aforementioned incentives to the experimental subjects prior to the second testing. The controls received a standard test administration that precluded such encouragement. Benton's investigation thus differed from those of Hurlock both in terms of sample size and instruments. Benton reported no significant differences between the gain scores of the praise and control groups. It should be noted in this context that the sundry behaviors that constituted the incentive

procedure obscured what effects, if any, that praise might have contributed.

Verbal Praise during Individual Testing

Klugman (1944) appears to have been the first among the verbal praise investigators to employ individually administered cognitive tests as a dependent variable. He also departed from his predecessors in two other ways: by contrasting the efficacy of praise and monetary rewards, and by making reinforcement contingent upon correct responses. His subjects were 72 pupils in grades two through seven, matched by sex (37 males and 35 females), race (38 white and 34 black), and grade. Examinees sat for both forms of the 1937 Stanford-Binet under one of four conditions: (a) praise or money for both sessions; (b) praise during the first administration and money during the second session; (c) money during the first administration; (d) praise during the second administration. The author did not detail the actual statements which constituted praise. He did, however, indicate that children earned between 5 and 15 cents in the monetary reward treatment.

Klugman went on to observe that there were no significant differences between the four treatment groups. It should be clearly noted, however, that this study did not employ a control group, and that this

omission precluded the determination of the treatments' efficacy vis-à-vis a neutral administration of the Binet.

After Klugman's study, research relating to the effects of praise on intelligence test performance seems to have been set aside for approximately 20 years. The first contemporary report to deal with the effects of verbal praise on the intellectual performance of students was a study by Tiber and Kennedy in 1964. Examined in this context, Tiber and Kennedy matched 480 second and third graders (gender of pupils was not mentioned) according to race and socio-economic level (i.e., equal numbers of lower class white and black children, and middle class whites), and assigned them to one of four groups (i.e., verbal praise, verbal reproof, candy reward, or control). The authors omitted operational definitions for these reinforcers. In addition, the experimenters did not use a pretest to establish the equivalence of the groups. The Stanford-Binet Form L-M was administered with the prescribed incentives given at the end of each subtest. Analyses revealed no main or interaction effects. Tiber and Kennedy concluded that "explanations of IQ differences between cultural groups must be based on

causes other than lack of intrinsic motivation provided by the intelligence test itself" (p.187).

A more stringent design was used by Bornstein (1968) in his assessment of the effects of verbal approval, disapproval, and neutral conditions on intelligence test scores. Bornstein pre-experimentally equated groups through a matched randomization procedure. Using the California Test of Mental Maturity as a covariate, Bornstein assigned 90 third, fourth, and fifth graders (with approximately equal representation of girls and boys) to one of the aforementioned conditions. A male examiner administered the Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding subtests of the 1949 Wechsler Intelligence Scale for Children. Approval was indicated by the following statements: "Good!," "Fine!," and "That was fine" (p.22). Disapproval consisted of comments like, "I thought you could do better than that," or "That wasn't too good" (p.22). In the neutral treatment, subjects were given no feedback regarding their performance. Noncontingent reinforcement was given after each examinee's response to the first item in each subtest and between subtests. Bornstein observed that the mean scores of subjects in the verbal approval group significantly exceeded the scores of the

subjects in the disapproval group on four subtests (Picture Completion, Picture Arrangement, Block Design, and Object Assembly) and on total Performance. Furthermore, the mean scores of children in the approval groups were statistically greater than the mean scores of children in the neutral administration group on all but the Picture Arrangement and Coding Subtests. Some sex differences were reported. On the average, boys in the verbal approval group attained significantly higher scores than boys in the disapproval group. Girls who were verbally praised, on the other hand, achieved significantly higher scores than girls in the disapproval and the neutral groups. Bornstein concluded that if these results were accurate, then changes in administration procedures to optimize test performance should take place. Such modifications, he added, would necessitate renorming the WISC.

Similarly, Witmer, Bornstein, and Dunham (1971) used the methodology and the subject characteristics (48 male and 42 female third and fourth graders) of Bornstein's 1968 research. They, however, used the Verbal (Arithmetic and Digit Span) as well as Performance (Picture Arrangement and Block Design) subtests of the Wechsler Intelligence Scale for Children (WISC) as dependent variables. Again, subjects in the

verbal approval treatment attained significantly higher mean standard scores than their counterparts in the disapproval and neutral groups. The researchers deduced that examiner verbal comments may have positive effects in real situations, but cautioned against assuming that the magnitude of improvement in scores would be comparable to the increases attained in their study. More specifically, they indicated that "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" (p.355).

The impact of material, social, and no incentives upon cognitive test performance of young, black middle class and lower class black children, was studied by Graham (1970/1971). The subject pool consisted of 96 nursery and kindergarten children with equal numbers of boys and girls from each income level. The 1959 edition of Dunn's Peabody Picture Vocabulary Test served as the covariate to the outcome measures, that is, the estimated IQ and the subtest scores (Information, Animal House, Sentence Memory, and Picture Completion) of the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1967). Children were randomly assigned to one of the three treatment groups. The praise condition

consisted of the following statements: "That's good," "That's very good," "That's fine," and "Very good" (p.22). Children in the reward incentive condition received candy. The no reward treatment was a control condition. The incentives were contingent upon correct responses; an unspecified partial reinforcement schedule was utilized. The three male and two female examiners, all black, were undergraduates or holders of baccalaureates from diverse disciplines. The testers administered the measures to children of both sexes.

Data analysis revealed no main effect for type of incentive, social class, sex of examiner, or sex of child. An effect of treatment by sex of child and sex of examiner on the Animal House Subtest was reported. In comparison with female examiners, male examiners elicited significantly higher scores from girls and boys in the neutral condition. In accounting for the nonsignificant differences of the experimental manipulations, Graham speculated that the individualized attention experienced by children in all treatments might have facilitated optimal motivation; hence, candy and praise would have negligible impact.

Bergan, McManis, and Melchert (1971) investigated the effects of social reinforcement (i.e., verbal praise), token reinforcement, and standard test

administration on the WISC Block Design performance of white, fourth grade boys and girls (24 of each sex). All subjects were initially tested according to the standardized directions. On the basis of these scores, the children were assigned to one of the three treatment groups and were subsequently retested. The praise condition involved the examiner exclamations of "Good," "Fine," "Right," "Very good," "Excellent," "Okay," and "Very good (p.874)," that were contingent on correct responses. Upon successful completion of an entire design, the examiner remarked, "I'll bet you can do the next one just as well" (p.874). In the token reinforcement treatment colored chips (which were exchangeable for money) were awarded for partial and total successes. The neutral condition involved the administration of the test according to the standardized WISC protocol.

Two outcome variables, accuracy and speed, were investigated. Total points for correct solutions comprised the accuracy measure. Speed, on the other hand, was defined as the absolute time score. Differences in the scores from pretest to posttest were analyzed. The accuracy difference scores of boys in the token reinforcement condition were significantly higher than the scores of boys in either the social

reinforcement or the control groups. Girls in the social reinforcement condition, however, attained significantly higher scores than girls in either the token reinforcement or the control group. Boys in the social reinforcement condition took significantly less time to complete the designs upon retesting than boys in either the token reinforcement or the control group. A sex by treatment interaction for speed was also noted. In the posttest, boys in the social reinforcement condition performed at a significantly faster speed than girls in the same treatment.

Unlike most researchers who sought support or disconfirmation of the meliorative effects of positive reinforcement during individual assessment, Bergan et al. (1971) interpreted the findings as clear evidence that extrinsic reinforcement may seriously bias performance. They argued that the results were in accordance with Wechsler's position (1958) that praise had a variable effect on children and therefore should be avoided (a stance that he modified somewhat with the 1974 publication of the WISC-R.)

Feldman and Sullivan (1971) examined the effects of "enhanced performance" and standard rapport on the WISC scores of 72 elementary school children. The subjects were matched by grade (not specified), sex (numbers of

each were not reported), and scores on the Otis School Ability Test. The children were subsequently assigned to the enhanced or to the standard rapport condition. The authors omitted operational definitions of the independent variables. Instead, they stated that "the only differences between the rapport conditions consisted of the amount of friendly conversation prior to and during the WISC testing and the inclusion of verbal reinforcement for the first correct response in each WISC subtest" (p.302). The content of the short form of the WISC was not reported. The authors indicated that subjects in the enhanced rapport treatment had significantly higher scores than their peers in the standard rapport conditons. In addition, it was reported that enhanced rapport was more effective for older children. The age or grade where the statistically significant effect was discerned was not disclosed. Feldman and Sullivan subsequently concluded that "the effect of rapport enhancement in raising obtained IQs apparently is attributable to the potency of E's friendliness and reinforcements in either directly increasing verbal productivity or in promoting greater verbal productivity despite the occurrence of other components of an anxiety response" (p.302). In retrospect, it should be pointed out that the authors'

failure to provide operational descriptions of the independent variables precludes the possibility of formulating a viable conclusion about the effects of verbal praise.

Racial (black and white) differences and social class (middle and lower) differences under contingent praise, contingent material reward, and traditional (i.e., standardized directions) conditions were studied by Wienges in 1971. Her criterion for the selection of subjects (120 parochial school students, divided equally according to sex, race, and social class in grades 4 through 6) was a score of 90 or higher on the Otis Quick Scoring Mental Ability Test (Forms Alpha and Beta). The children were randomly assigned to one of the three treatments (i.e., verbal praise, material reinforcement, and traditional test administration) as they sat for an abbreviated form of the WISC (i.e., the Doppelt). Subjects in the verbal praise treatment were commended for correct responses with the phrases, "That's good," "You're doing well," or "Right" (p.27). Subjects in the material reinforcement condition, on the other hand, received candy, charms, or baseball cards for correct responses. Children in the third group, standardized administration condition, received neither verbal nor material rewards. Differences between the Otis and the

WISC IQ scores constituted the outcome variable. Wienges found no main effects for sex, grade, race, and type of reinforcement on the selected measures. She did, however, report differential effects. Lower social class black pupils made significantly greater gains than white middle class students in the material reward condition; white lower class children evinced significantly greater gains in the material reward condition than they did in the verbal reward condition. Black and white children from low socio-economic backgrounds made significantly greater gains than white middle class students in the material reward condition. In addition, black lower class subjects improved more than black and white middle class subjects in the traditional condition. Further, lower class white subjects scored significantly better than white middle class subjects in the traditional condition. In turn, black middle class pupils evinced statistically greater gains than white middle class pupils in the traditional condition. The differences that Wienges noted in the control groups led her to theorize that the individualized attention itself may be positively reinforcing, an experience that black and low income students may not often encounter. Wienges's overall

conclusion, however, was that material rewards were more effective.

The influence of incentive conditions on the intellectual performance of black children was also investigated by Quay (1971). Taken in this context, 100 three- and four- year old Head Start pupils (55 boys and 45 girls) were randomly assigned to one of four treatment groups and given the Stanford-Binet (Form L-M) by black examiners. No pretest of intellectual functioning was administered which, if used, would have ensured that the groups were equivalent in terms of cognitive functioning prior to treatment. Two levels of reinforcement (praise and candy) and two levels of test language (standard English and black English) comprised the independent variables. The incentive conditions were contingent upon correct responses. In the verbal praise condition, the examiner said, "That's good." The preschoolers in the candy condition were unaware that the sweets were awarded on a contingency basis. In the standard English treatment, the examiner read verbatim statements from the test manual; no feedback was given. In the black English treatment, the examiner presented a black language protocol. An example of the latter is the following direction for Pictorial Similarities and Differences at year 4-6: "See all dese crosses? See how

mos' of 'em de same? Here go one (pointing) what ain't like de uvvers. Put your finger on de one what ain't de same like de uvvers" (p.8).

Quay observed no main or interaction effects. She noted that the IQ scores for her sample were close to the normative mean and higher than those of children with similar demographic characteristics. Quay went on to suggest that the subjects' extant level of measured intelligence might have obviated the need for external incentives. The absence of a pretest and the lack of a control group, however, weaken the validity of her conclusions.

In a related effort, Quay (1975) enlisted the participation of 92 (46 males and 46 females) black fourth graders. In this procedure, type of reward (verbal praise in standard English or monetary rewards) and sex of subject constituted the independent variables. Again, reinforcement was dependent upon presenting the correct responses to items of the Stanford-Binet. The verbal praise condition was identical to the treatment in Quay's 1971 experiment. Unlike the children in the earlier study, however, the subjects in the material reward condition were apprised of the contingency basis of the incentives. More specifically, they were told: "At certain times during

this test, if you have enough correct answers, you will get a nickel. The more correct answers you have, the more money you will make" (Quay, 1975, p. 133). Data analyses revealed a main effect for sex (i.e., the mean scores of boys in each treatment were higher than the scores of girls), but revealed no main effect for reward nor interaction effect. Quay concluded that the study failed to support Havighurst's 1970 theory that lower-class children perform better when given material rewards than when they are given symbolic (i.e., verbal praise) rewards. It is important to note in this context that this investigation, like Quay's 1971 study, lacked a control group which, if used, could have clarified whether either or both of the treatments were superior to standard (i.e., neutral) administration.

Galdieri, Barcikowski, and Witmer (1972) compared the performance of 72 (gender undisclosed) rural middle and low socio-economic white third graders during administration of the entire WISC under conditions of praise or neutral comments. Galdieri et al. did not check on the initial equivalence of their groups. Verbal praise consisted of three examiner statements. The first involved declaring "good" after the first response of each subtest. The second involved exclaiming "That was hard, wasn't it? But you are doing

good" after the first incorrect response per subtest. Finally, examinees were told "Let's go on to something else," or "Now try these," between subtests. The neutral group attempted items without the benefit of subsequent examiner comments. The authors observed no significant differences between the scores of the treatment groups, and it was concluded that "it would be unfortunate if we had to worry about test results because of the loquaciousness of the examiner" (p.408).

From the mid 1970s through the early 1980s several researchers (Miller, 1974; Saigh, 1981; Saigh & Payne, 1976, 1979; Tufano, 1976) attempted to verify Binet's and Simon's original assumption that praise facilitated the test scores of mentally retarded students. Miller (1974) compared the effects of three types of test administration (i.e., contingent verbal praise, contingent token reinforcement, and standard procedures) and two levels of social deprivation (i.e., relatively deprived and relatively undeprived) on raw scores of the Peabody Picture Vocabulary Test (Forms A and B). Sixty institutionalized retardates were matched for sex (30 males and 30 females), origin of retardation (organic or familial according to criteria established by the American Association for Mental Deficiency), age (ranging from 9 years to 21 years, 6 months) and IQ.

Gain scores on the PPVT comprised the outcome variable. The Social Interaction Inventory, apparently a form devised by the author for this investigation, was completed for each subject by four staff persons at the facility. Based on their ranks on this measure, subjects were divided into two groups: high interaction (relatively undeprived) and low interaction (relatively deprived or isolated).

At this point, the Peabody was administered. Examined in this context, subjects in the verbal praise condition received examiner praise, (e.g., "That's very good," "Hey, another one right," and "Right again; you must do very well in school") for correct responses. In addition, after every third correct response, the tester spent 15 to 30 seconds in conversation with the subject. For the token treatment, the examiner rewarded the subject with a penny for each correct response and the money was subsequently traded for candy. In the control condition, brief noncontingent approval statements were presented. Miller did not, however, describe these comments.

Miller observed no main effects on the raw or mental age change scores. He did report, however, an interaction effect. The relatively undeprived subjects (i.e., those with high SII scores) in the social

reinforcement condition attained significantly higher raw scores than the relatively deprived individuals. Miller hypothesized that a history of failure in social and academic situations similar to the PPVT testing made the experience aversive for the low SII subjects.

In 1976, Tufano matched 30 black and 30 white educably mentally retarded boys, aged 10 to 14 years, on the basis of previous IQ scores. The subjects were then randomly assigned to one of two treatment groups: contingent (effort) verbal praise and contingent (performance) verbal praise during administration of the WISC-R. The subjects in the effort condition had to exhibit one of the following behaviors in order to receive verbal praise: eye contact with the examiner, verbal attempt to answer the question, attempt to perform the task manually, or leaning forward. These behaviors were reinforced by the examiner, who used the following statements: "You are trying hard;" "You are working hard;" "You made a good try;" "I like your helping me;" "I'm so glad you enjoy learning." These comments were offered after the third effort in a subtest and then after every other effort. In the performance condition, the examiner reinforced each correct response or solution with one of the following remarks: "Your answer is correct;" "Your answer is

right;" "You gave a good answer;" "I like your doing a good job;" I'm glad you did a good job."

Tufano reported no main effect for treatment on the WISC-R Verbal, Performance, and Full Scale. The only differential effect for race and treatment was noted on the Information subtest where blacks in the effort treatment did better than whites. Moreover, whites in the performance condition achieved significantly higher scores than blacks in the performance condition. Tufano explained the results in terms of the law of diminishing returns: "Perhaps there is a critical number of verbal reinforcements beyond which such reinforcement loses its effectiveness" (p.84). He acknowledged that the small sample size (five subjects per cell) and nonverbal cues (e.g., testers' unwittingly smiling or nodding) during the course of the study might have affected the outcomes. Problematic in this study was the lack of a control group.

In lieu of eliminating nonverbal cues as Tufano attempted, Mayfield (1977) paired nonverbal behaviors (e.g., smiles and nods) with social reinforcement (noncontingent positive verbalizations) and compared the combination with the presence or absence of a prior relationship (defined as a fifteen minute "warm-up")

between the examiner and the subject. The subjects were 24 male (15 black and 9 white) and 24 female (11 black and 13 white) clients at a psycho-educational center. The children, who ranged in age from 7 through 11 years, were selected on the basis of cognitive ability (average range on the Slosson Intelligence Test) and extreme scores on the Quay-Peterson Behavior Problems Checklist (Quay, 1967). The gain scores on the WISC-R comprised the dependent variable. In the positive verbalization group, reinforcement was given after every third response. The comments in the examiner's repertoire included "Fine," "Great," "Good for you," and "Good trying." These remarks were made in tandem with smiling, nodding, and other nonverbal behaviors to underscore approval. In the non-reinforcement condition, the examiner made no comment during test administration. To avoid the child's inferring that no comment indicated poor performance, the tester offered positive remarks at the conclusion of the session. Mayfield reported that the mean scores of subjects who were reinforced significantly exceeded the scores of those who were not reinforced. In addition, children in the prior relationship-reinforcement group and children in the prior relationship-no reinforcement group achieved

significantly higher scores than those in the prior relationship without reinforcement group. Thus, the findings favored both social reinforcement and prior relationship. Mayfield conjectured that the remarks of the examiner might have provided the interest the child needed to limit other distractions.

In a series of three experiments (Saigh, 1981; Saigh & Payne, 1976, 1979) the effects of verbal praise on the Wechsler (WISC or WISC-R) performance of educably mentally retarded students was explored. The investigations were similar with respect to sample characteristics (black and white junior high school age students of both sexes); instruments (either the full WISC-R or a short form consisting of Arithmetic, Digit Span, Picture Completion, and Block Design); and use of noncontingent reinforcement. The independent variables were either verbal praise regardless of the correctness of the response, e.g., "Very good," "Keep it up," "That's the stuff," "I like the way you're working," for the experimental group, or neutral comments such as "Let's try this," "How about this," "Here is the next" were made to the control group subjects. The studies varied in terms of sample size, schedule of reinforcement, and additional treatment variables.

In the earliest study (Saigh & Payne 1976), examiners administered a short form of the WISC to 40 educably mentally retarded (EMR) students. Approximately equal numbers of black and white pupils of each sex constituted the sample. The examiners in the approval condition praised the subject after the first and second items in the WISC Arithmetic, Block Design, Picture, and Digit Span subtests. Praise was also provided between subtests. Neutral comments were made according to the same rate of reinforcement. Statistical analysis revealed an overall effect favoring the verbal praise procedure. Specifically, the verbal praise group obtained significantly greater scores on the Block Design and Digit Span subtests. On the other hand, significant differences were not noted on the Arithmetic and Picture Completion subtests. Saigh and Payne suggested that their results supported the efficacy for positive verbal comment procedure as a variable for facilitating the performance of educably mentally retarded students during standardized testing sessions. They tempered this conclusion with a recommendation for exploring the effects of verbal reinforcement on the Full Scale Score of the WISC-R.

In the second study, Saigh and Payne (1979) examined the effects of incentive conditions (verbal

praise versus token reinforcement) and schedule of reinforcement (fixed versus continuous) on the WISC-R Block Design, Digit Span, Arithmetic, and Picture Completion performance of 120 (60 male and 60 female) educably mentally retarded students. In doing so, subjects were randomly assigned to one of six conditions. In the first condition (i.e., fixed ratio-verbal praise), subjects were verbally praised, "That was very good; keep up the effort" after the first, second, and third item on each subtest without regard to correctness and between subtests. In the second condition (i.e., continuous verbal praise) subjects received the same statements as their counterparts in the first group. Verbal praise, however, was provided after every response and between subtests. In the third condition (fixed ratio-token reinforcement), examiners gave tokens that were exchangeable for chocolate after the first three responses of a subtest and between subtests. In the fourth condition (continuous token reinforcement), subjects were given exchangeable tokens after every response and between subtests. The fifth condition (fixed ratio-neutral) involved the use of neutral non-evaluative examiner comments (e.g., "Let's try this one") after the first, second, and third examinee

response and between subtests. The sixth condition entailed a continuous neutral statement regimen wherein neutral comments were made after every response and between subtests.

Data analysis revealed that the Digit Span, Arithmetic, and Picture Completion scaled scores of the verbal praise and the token reinforcement groups significantly exceeded the scaled scores of the controls. No significant differences were observed as a function of the type of reinforcement schedule that was used. Moreover, no significant interaction effects were noted. Saigh and Payne subsequently indicated that the differential treatment effects might have been due to the nature of the mental operations that the various WISC-R subtests reflect. More specifically, it was indicated that unlike the Arithmetic, Digit Span, and Picture Completion subtests, Block Design involves analysis, synthesis, and reproduction of abstract figures. With this in mind, the authors reasoned that the educably mentally retarded students might have had an especially difficult time on this complex task due to the nature of their impairment and that effort alone could not induce successful completion of the items.

Citing how the WISC-R Full Scale IQ (i.e., the aggregate of 12 subtests) is a critical factor in

placement decisions of special education students, Saigh (1981) investigated the effects of positive reinforcement during administration of the Full Scale WISC-R. In this investigation, 22 black and 18 white EMR students (with equal representation of both sexes) comprised the sample. The subjects were randomly assigned to either an experimental group (verbal praise) or to a control group. Verbal praise (e.g., "very good") was given after the initial four items, after every other response beyond the fourth response, and between subtests. In the control group, the examiner made comments such as, "Let's try this one," at the same rate of reinforcement as in the verbal praise group. Saigh reported that praise had a significant impact on the Verbal, Performance, and Full Scale IQ. He also noted that the Vocabulary, Arithmetic, Picture Completion, Digit Span, and Coding scaled scores of the verbally praised subjects significantly exceeded the scores of the controls. Saigh suggested that the Vocabulary and Arithmetic subtests might have been especially sensitive to the experimental manipulation because basic competencies in these areas were emphasized in the curriculum of the EMR students. Saigh and Payne (1976) reported that in their study, verbal praise did not appreciably increase the scores in

Arithmetic and Picture Completion for EMR subjects. One may speculate that differences in duration of the experiment (i.e., Full Scale administration versus the administration of four subtests), in examinee variables (e.g., the pre-experimental IQ ranges of the two studies differed), and between the original and revised Wechsler instruments might have accounted for the discrepant findings.

In 1977 Goh and Lund compared the effects of verbal feedback, noncontingent praise, and standard administration on the intellectual performance of preschool children. The subject pool of 90 was balanced according to gender and socio-economic status. (The racial composition of the sample was not disclosed.) All subjects were initially examined with the Peabody Picture Vocabulary Test (Dunn, 1955) according to manual directions. Subsequently, the subjects were randomly assigned to one of three groups. Each group received a different administration of the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1967). More specifically, the first group received standard administration with no verbal reinforcement. The second group received noncontingent examiner praise such as, "Good," "Very good," or "You're pretty smart" (p.1012). In the third group, the examiner reinforced responses,

regardless of correctness, by such comments as, "Right," or "That's a good answer." The authors did not specify the schedule of reinforcement. The Peabody scores served as covariates for the Wechsler scores. The authors reported no significant main effect for type of examiner comment. In addition, no interaction effect (treatment by social class membership) attained statistical significance. Goh and Lund discussed two possibilities for the lack of significant results. First, excessive verbal feedback might have served as a distractor. Second, the preschoolers might have been too young to be affected by statements such as "Good" or "Right." According to the investigators, these reinforcements are usually associated with a higher level of cognitive development. Goh and Lund concluded that "Although praising a child's efforts contributes greatly to satisfactory rapport in a testing situation, test users should not assume an unqualified positive linear relationship between kind of verbal reinforcement and IQ" (p.1013).

The type of dialect (standard English and black English) which Quay (1971) investigated re-emerged as an independent variable in a series of investigations (Terrell, Taylor, & Terrell, 1978; Terrell, Terrell, & Taylor, 1980, 1981). In the first experiment (Terrell

et al., 1978), 80 lower socio-economic black second graders (the gender composition was not disclosed) were randomly assigned to one of the four treatments and were tested with a short WISC-R. The senior author, a black doctoral level psychologist, was the examiner. Subjects in the non-reinforcement group were tested according to manual procedures. Children in the candy reward condition were reinforced with an M&M after each correct response. Subjects in the traditional social reinforcement group were rewarded for their correct responses by such remarks as "Good" and "Fine." Children in the culturally relevant social reinforcement group, on the other hand, were congratulated for correct responses with such praise as "Good job, blood," and "Nice job, little brother" (p.1539). The report did not specify what constituted the modified test; also absent was a description of attempts to ensure that the four groups were initially alike. The results disclosed no significant differences between the control and the traditional social rewards groups, nor between the social reward and the candy reward groups. The scores of subjects in the candy reward condition, however, significantly exceeded the scores of the children in the control condition. Finally, the culturally relevant social praise group scores exceeded significantly the

scores of the control and traditional social reinforcement groups. The researchers subsequently deduced that the use of appropriate, i.e., culturally relevant social reinforcers, is more effective than tangible reinforcers with black children during cognitive testing.

In a subsequent study, Terrell, Terrell, & Taylor (1980) again contrasted nonreinforcement (standard administration), tangible reinforcement (M&M candy reward), traditional reinforcement (verbal comments such as "Good" and "Fine"), and culturally relevant reinforcement (remarks such as "Nice job, little brother," and "Good work, young soul") with 120 black males, aged 9-11 years old. Reinforcement was contingent on correct responses. A second independent variable (race of examiner), was added. The investigators did not provide information about the pre-experimental equivalence of the groups or the composition of the short WISC-R. No significant main effect for race of examiner was detected. Terrell et al., however, noted a main effect for type of reinforcer: children given tangible rewards (i.e., candy) obtained significantly higher scores than children given no reinforcement or given the traditional social reinforcement. Two interaction effects were also

reported. Children in the culturally relevant reward condition with a black examiner obtained significantly higher scores than children in the same experimental situation with a white examiner; subjects who were tangibly rewarded by white examiners performed significantly better than their peers in other experimental conditions with white examiners. Terrell et al. concluded that the results suggested that tangible reward was the most effective reinforcer for white examiners to use with the black children, whereas tangible and socially relevant reinforcers were equally efficacious when administered by black examiners to black children.

The third study, by Terrell, Terrell, and Taylor (1981), involved 100 mildly retarded black male pupils whose ages ranged from 9 to 11 years. In this instance, IQ (WISC-R or Stanford-Binet) scores used for special education placement served as the pretest scores. Subjects were randomly assigned to one of four experimental conditions in which they subsequently received a different administration of the entire WISC-R. The children received either no reinforcement, a candy reward, traditional social reinforcement (e.g., "Good" and "Fine"), or culturally relevant social reinforcement (e.g., "Good work, young soul," and "Nice

job, little brother"). Data analysis revealed that the culturally relevant social reinforcement and the tangible reward treatments were equally effective in terms of raising test performance in comparison with nonreinforcement and traditional social reinforcement. The investigators concluded that the type of reinforcer has an important effect on IQ scores of black pupils labeled as mentally retarded. Moreover, the authors observed that the gain scores of the subjects in the culturally relevant social reinforcement and in the tangible reward groups were higher than the cutoff score commonly accepted for diagnosis of mental retardation. These findings prompted the researchers to question the validity of the original diagnosis of the selected sample.

Summary of Effects of Praise during Individual Testing.

Examined in toto, 24 studies were reviewed. Of these, three reports (Benton, 1936; Hurlock, 1924, 1925) were included because of their historical importance in introducing verbal praise as an independent variable in cognitive assessment. Of the remaining 21, 14 indicated that praise facilitated performance. Six investigations failed to evince a significant effect. The majority of the studies (both in favor of treatment effects as well as those that did not support an appreciable effect for

verbal praise), however, had one or more methodological weaknesses. Four studies, (Klugman, 1944; Quay, 1971, 1975; Tufano, 1976) did not use a control group. Analogously, in the Feldman and Sullivan study (1971) verbal praise was embedded in an enhanced rapport situation so that its utility could not be separated from the overall effect of rapport. Seven studies (Galdieri et al., 1972; Quay, 1971, 1975; Terrell et al., 1978; Terrell et al., 1980; Tiber & Kennedy, 1964; Wienges, 1971) did not provide sufficient information regarding the initial equivalence of the groups. Also problematic were nine investigations (Bergan et al., 1971; Graham, 1971; Klugman, 1944; Miller, 1974; Quay, 1971, 1975; Terrell et al., 1978; Terrell et al., 1980, 1981) that used contingent reinforcement for correct responses or solutions. This procedure is contrary to the recommendations of test developers (Kaufman & Kaufman, 1983a; Terman & Merrill, 1973; Wechsler, 1974) who emphasize that effort, rather than achievement, be rewarded. The technique is also questionable from a psychometric point of view because the resulting scores were obtained in a manner radically different from the standardization procedures. Hence, comparisons with test norms may be invalid.

Of the original 24 studies, seven (Bornstein, 1968; Goh & Lund, 1977, Mayfield, 1977, Saigh, 1981; Saigh & Payne, 1976, 1979; Witmer et al., 1971) avoided the preceding pitfalls. Of these, all but the research of Goh and Lund concluded that verbal praise improved test performance. Praise had a positive effect upon diverse groups: elementary school children, institutionalized retarded children, and mildly retarded junior high school aged students.

Despite evidence attesting to verbal praise as an enhancer of motivation during cognitive testing, psychologists may not want to rely exclusively on this technique or to assume that it is beneficial with other types of children such as low achievers who comprise the majority of test referrals (c.f. Sattler, 1988). For example, some students may find verbal comments to be distractors or sources of embarrassment (Brophy, 1981, 1983, 1986). Other students may infer from verbal praise that they lack ability (Pearl, 1985). In addition, some researchers (Brophy & Everston, 1976; Stallings, 1975) report that although the frequency of praise may have a positive correlation with achievement, these associations are often weak. In such cases, praise may be contraindicated. An alternative technique, material reward, is also problematic,

especially on a contingency reinforcement schedule, because the obtained results cannot be compared with the normative sample. Thus, other avenues for improving motivation must be sought.

Social Learning Theory

Theoretical Underpinnings

Peer modeling is a relatively untried method for improving examinee performance during the administration of standardized testing. This technique is esconced within the framework of Bandura's (1977, 1986) social learning theory. Bandura posits that most learning occurs vicariously, that is, through observations of others who serve as models. According to this paradigm, the observer does not need to be reinforced directly in order to learn. Viewed in this context, Bandura (1977) suggests that "reinforcement is considered a facilitative rather than a necessary condition because factors other than response consequences can influence what people attend to" (p.37). His conceptualization of learning thus departs from Skinner's tripartite operant conditioning paradigm (1953) of stimulus, response, and a reinforcing stimulus (e.g., the scheme upon which the research on verbal praise in mental testing is based.)

Bandura (1977, 1986) asserts that modeling may result in new behavior (observational learning), the

strengthening or weakening of inhibitions (inhibition-disinhibition) or the execution of behaviors already in the individual's repertoire via prompting (response facilitation). The last effect, response facilitation, is germane to the school psychologist's attempt to elicit the best performance of the child during cognitive assessment.

In response facilitation, Bandura (1977) posits that:

> The actions of others can also serve as social cues for eliciting preexisting behavior. Response facilitation is distinguished from observational learning in that nothing new is learned, and from disinhibition, because the behavior is socially acceptable and therefore is unencumbered by restraints. In response facilitation, the modeled actions function simply as social prompts (p.40).

Model-Observer similarity. Also central to social learning theory in general and response facilitation in particular is the assumption that the observer's perceived similarity to the model will increase the liklihood that the observer will emulate the modeled behavior. Among the personal attributes purported to heighten identification are age, sex, and competence (Akamatsu & Thelen, 1974; Bandura, 1977, 1986; Hartup & Lougee, 1975; Rosenthal & Zimmerman, 1978; Schunk, 1987).

The association between model-observer similarity and observer behavior, however, is not perfectly linear and is qualified by situational factors. Schunk (1987), for example, concluded from the literature that peers may be more influential than adults in those instances when the observer is unsure about his or her own capability. The research on model sex is likewise equivocal. Sex-linked modeling appears to be beneficial in situations that are ambiguous as to the gender appropriateness of the task (Bussey & Bandura, 1984; Perry & Bussey, 1979). In such circumstances, observers are more likely to take their cues from models of the same sex. With increased age, however, greater flexibility in cross sex behavior may become apparent (Bussey & Bandura, 1984; Perry & Bussey, 1979).

Competence can be examined from diverse perspectives (i.e., model and observer; the relative competence of either; and the use of mastery or coping model). Schunk (1987) observed that empirical research supports the thesis that children emulate competent rather than incompetent peers. He also noted that "when information on competence and age do not match, children are more inclined to model equally competent but younger children rather than age-mates of lower competence than themselves (p.167)." Low competent observers are more

likely than high competent observers to evince such imitation (Akamatsu & Thelen, 1974).

Schunk, Hanson, and Cox (1987) distinguish between coping and mastery modeling: "Coping models initially demonstrate the typical fears and deficiencies of observers but gradually improve their performance and help gain self confidence, whereas mastery models demonstrate faultless performance from the outset (p.54)."

The mastery or "fearless" model, as applied to therapeutic situations (e.g., reduction of phobias or test anxiety) was predicated on the hypothesis that a calm model would be less emotionally threatening to the client than a nervous model; the fearful model was presumed to arouse negative affect in the client which in turn would interfere with subsequent performance (Bandura, 1968). This view is consonant with systematic desensitization theory but varies with implosive therapy which maintains that high levels of arousal (e.g., as experienced by the viewer of a coping model) affect behavioral change (Jaffe & Carlson, 1972). The preponderance of evidence from comparisons between coping and mastery procedures suggests that coping techniques are as effective as or superior to mastery models. These effects were demonstrated in clinical

situations such as avoidance behavior and pain reduction (Klorman, Hilpert, Michael, LaGana, & Sveen, 1980; Kornhaber & Schroder, 1975; Vernon, 1974) as well as in academic achievement activities (Schunk & Hanson, 1985; Schunk et al., 1987).

Multiple models may increase the likelihood that the observer can identify points of similarity between himself or herself and one of the models (Bandura, 1969; Schunk, 1987; Thelen, Fry, Fehrenbach, & Frautschi, 1979).

Symbolic Modeling

Vicarious experiences can take two forms: live modeling or symbolic modeling in the form of films or videotapes (Thelen et al., 1979). The second method offers several advantages to practitioners: the opportunity to capture naturalistic sequences that would be difficult to replicate in a clinical setting; greater control over the composition of the modeling scene; repetition of the observation; use of multiple models who might not be as readily available in live situations; reuse of the materials; (cf. Thelen et al., 1979).

Use of Narration during Symbolic Modeling

Thelen et al. (1979) further point out that symbolic modeling also permits incorporation of

narration that may enhance the four subprocesses in modeling identified by Bandura (1977): attention (i.e., the observer's perception or awareness of relevant environmental aspects of the behavior in question); retention (i.e., the storage, representation, and retrieval of verbal and imaginal information in memory); production (i.e., the subsequent display of such learning through overt behavior); and motivation (i.e., the inducement to perform the observed behavior.) Thelen et al. (1979) suggest that "narration should facilitate attention to the model and verbal labeling of the critical behavior and thereby increase the effectiveness of modeling interventions" (p. 712).

While the experimental evidence suggests the efficacy of symbolic modeling alone, the incremental effects of narration and symbolic modeling have not been widely investigated (Thelen et al, 1979). Evidence from cognate research on verbal instructions and modeling, however, indirectly supports the proposition that narration in tandem with modeling may produce effects over and above symbolic modeling alone. For example, Zimmerman and Rosenthal (1974) noted in their extensive literature review on children's observational learning of rule-governed behavior that verbal instructions enhanced observational learning. They concluded, "When

a modeling display is accompanied by symbolic codes (from the model or experimenter) which provide a mnemonic summary of the rule governed response, concept attainment and retention are enhanced" (p.38).

Effects of Symbolic Modeling on Individual Cognitive Testing

Although the benefits that can be derived from symbolic modeling are well established in the treatment literature, its application during the assessment of intellectual ability has been limited. Recent computer and manual searches of the literature yielded only three pertinent studies (Jaffe & Carlson, 1972; Piersel, Brody, & Kratochwill, 1977; Raskind & Nagle, 1980). Added to this modicum of research is a recent unpublished report (Pollock, 1988). Two of the investigations (Jaffe & Carlson, 1972; Raskind & Nagle, 1980) examined the impact of modeling on diminuition of test anxiety while the most recent study (Pollock, 1988) approached modeling from the perspective of response facilitation. One study (Piersel et al., 1977) viewed test apprehension from a motivational perspective.

Jaffe and Carlson (1972) investigated whether type of model (calm or anxious) and type of feedback (positive or negative) would affect cognitive test performance of college students. Twenty female and ten

male undergraduates, judged test anxious by their scores on Sarason's (1971) Test Anxiety Questionnaire (TAQ), initially sat for a test composed of items from selected subtests of the 1955 Wechsler Intelligence Scale for Adults (WAIS), the WISC, and newspaper anagrams. The subjects were then assigned on the basis of their WAIS scores to one of four experimental conditions (Videotape A, B, C, or D) or to a control group which viewed no tape.

Each subject viewed one of four videotapes. The videotapes were similar with respect to duration (one hour), actors (a male and a female subject with an examiner), and test content (WAIS Vocabulary, Object Assembly, and Digit Span). The behaviors of the model and the examiner varied from tape to tape.

In Tape A (Calm-Positive) the models were very calm and business like and demonstrated clear organization of thought; the examiner gave positive feedback and remarked at the end of the session that the person had performed well above his or her age norms. In addition, the examiner predicted a successful future for the examinee. Tape B (Anxious-Positive) featured the same scenario as Tape A. In this case, however, the subjects were clearly anxious, (i.e., fidgety, restless, and overly concerned with their performance). In Tape C

(Calm-Negative) the actors maintained a calm demeanor but performed badly and were given negative feedback at the end of each subtest. Moreover, the examiner told the examinees that they were functioning well below average and that it was rather incredible that they were college students. Tape D (Anxious-Negative) showed the anxious subjects performing poorly and experiencing the same consequences as in (C). Subjects in the control group did not observe a videotape.

Prior to the experimental manipulation, subjects were advised to pay close attention to the videotapes because they were going to rate the models on levels of anxiety and performance. This procedure was implemented to ensure that the salient features were perceived as planned. After completing the rating sheet, the experimental subjects were administered a parallel form of the pretest. The investigators reported that the Calm-Negative, Anxious-Positive, and Anxious-Negative videotapes were associated with increases in the post-test scores in comparison to the the control group. There was no significant difference between the scores of participants in the Calm-Positive and control conditions. The effects, however, were limited to the subtests that were attempted by both the subjects and the videotaped models. In other words, the treatment

effect did not generalize to the subtests that were excluded from the videotape (Anagrams, WAIS Coding and Picture Arrangement subtests). Jaffe and Carlson concluded that the results gave qualified support to their hypothesis that exposures to anxious models or negative modeling consequences may improve the performance of test anxious individuals more than exposures to the fearless (low arousing) models as advocated by Bandura (1969).

Piersel et al. (1977) compared correctness of response feedback, vicarious modeling, and standard test administration on the WISC-R scores of children from black, Mexican-American, and ethnically mixed families. The sample of 32 boys and 31 girls ranged in age from 8 to 10 years. The students were randomly assigned to one of three procedures. In this context, they sat for the WISC-R according to the adaptations proposed by Silverstein (1968). In the first procedure (self-monitoring feedback), children were informed about the correctness of their responses and were awarded points for correctness. The children recorded their points on a record form on which they could compare scores with the total number of points possible for each subtest. In the second group (the vicarious modeling treatment), subjects watched a 7 minute videotape of a

white female examiner who made noncontingent positive statements such as "Very good," and "You're doing great," to a minority female student who was taking a test similar in content to the WISC-R. The interaction was spontaneous and the child (i.e., model) succeeded on approximately 60% of the items. According to the experimenters, viewers could readily see that the model was sometimes wrong and that no negative consequences were associated with incorrect responses. The third group sat for the WISC-R under standard testing conditions (i.e., they did not observe a modeling tape). It was observed that the pretest vicarious modeling was statistically more effective than either the standard administration or the feedback experience on the Full Scale WISC-R scores. Piersel et al. concluded that modeling is a viable technique for facilitating the performance of students during administration of the WISC-R. The researchers suggested that modeling may be a form of desensitization that lowers evaluation apprehension.

Raskind and Nagle compared the effects of three film presentations (i.e., a reenactment of a testing session; a similar film that incorporated Meichenbaum's didactic coping method [1971]; and an unrelated film) on selected WISC-R subtest performance of elementary school

children. Ninety-six white fifth graders (48 boys and 48 girls) initially marked the Otis-Lennon Mental Ability Test (OLMAT), Elementary Level II, Form J. In addition, subjects completed the Test Anxiety Scale for Children (TASC) (Sarason, Davidson, Lightall, Waite, & Roebush, 1960). Pupils whose scores fell in the lowest quartile of the TASC were regarded as low test anxious (LTA). Subjects in the highest quartile were considered high test anxious (HTA). Subjects who scored in the mid range were deemed middle test anxious (MTA). Subjects were matched by sex, OLMAT scores, and anxiety level. They were subsequently randomly assigned to one of three experimental groups. Children in each treatment group individually viewed a different 10 minute film. The first film (observation film) depicted an anxious white girl and white boy who attempted WISC-R -like items that were administered by a white female examiner. It is important to note in this context that the examiner provided verbal encouragement. The second film (coping treatment) featured the same models. In this film, however, the examiner provided encouragement and strategies for test taking. The third film (unrelated film) depicted a class on a field trip. After the presentation of each film, the WISC-R Information, Arithmetic, Vocabulary, Digit Span, Picture Completion,

Block Design, and Coding subtests were administered. A 3 (low, moderate, or high level of anxiety) by 2 (male or female) analysis of covariance failed to detect any main or interaction effects. Raskind and Nagle accounted for the disparity between their results and the previous research on the bases of a more stringent experimental design which reduced examiner bias and differences in subject characteristics. Elsewhere, Raskind (1978) conjectured that the unrelated film might have had a relaxing effect on the subjects. According to this rationale, the film might have constituted an experimental rather than a control condition.

Although the Raskind and Nagle investigation overcame the problems that were associated with the Jaffe and Carlson (1972) and Piersel et al. (1977) studies, their own design and statistical analysis were problematic. Although their sample contained more subjects, the selected subjects were blocked into 18 cells (with an n of 4 or 8 per cell). The limited number of cases per cell decreases the likelihood of observing a significant treatment effect. Finally, the variances were so extreme (i.e., a violation of homogeneity of variance) that a non-parametric analysis might have been more appropriate.

Most recently, Pollock (1988) compared the effects of three treatment conditions (i.e.,mastery model videotape with standard test administration; unrelated videotape with examiner verbal praise; and unrelated videotape with standard administration) on the WISC-R performance of elementary school children. Fifty-three (30 boys and 20 girls) fourth and fifth graders of an urban parochial school sat for the Otis-Lennon School Ability Test (Otis & Lennon, 1982). The students were subsequently matched on the basis of sex and grade and randomly assigned to one of three treatment conditions. Participants in the first group (Mastery Modeling - Standard Administration) individually watched a five minute rehearsed videotape that featured a same sex model successfully complete several items from the Gestalt Closure and Matrix Analogies Tests of the Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983a). Halfway through the videotape, the examiner stopped the tape. The tape was subsequently rewound and replayed. During the replay, the examiner pointed out how the model was continuing to take his or her time and was not getting discouraged despite the increasing difficulty of the items that were presented. After viewing the remainder of the tape, the examiner administered the Picture Completion, Arithmetic, Block

Design, and Digit Span subtests according to the standardized WISC-R regimen. Children in the second group (Unrelated Tape and Verbal Praise) watched an unrelated commercial videotape of a student who was demonstrating a math trick to a companion. In the subsequent WISC-R administration, the examiner offered noncontingent verbal praise (e.g., "Good," "Fine," "That was very good") after the first, second, and third items in every subtest, after every third response thereafter, and between subtests. Pupils in the Unrelated Videotape - Standard Administration group saw the same videotape as the students in the Verbal Praise group. When they sat for the WISC-R, however, the examiner presented the same neutral, non-evaluative comments that were given to subjects in the Mastery Model condition. Statistical analyses did not reveal a statistically significant main effect or interaction effects. The author subsequently suggested several possibilities for the lack of effect: the small sample size; relatively brief exposure to the independent variables; use of mastery rather than coping models; and the possibility that the subjects did not identify with the model.

Summary of Effects of Symbolic Modeling upon Cognitive Test Performance

Empirical studies have documented the relative efficacy of verbal praise on intelligence test performance. Findings from the limited research on the effects of modeling on cognitive test performance have, however, been less consistent. The results of the earlier investigations (Jaffe & Carlson, 1972; Piersel et al., 1977) were promising. Later studies (Pollock, 1988; Raskind & Nagle, 1980) found that modeling had no significant impact on intelligence test performance.

The equivocal results may be a function of sample characteristics, methodological weaknesses, and diverse procedures. The study by Jaffe and Carlson (1972) found that videotapes of anxious examinees and videotapes of negative consequences of test completion significantly affected test scores. This study, though well designed, was conducted with college students. The utility of modeling experiences with a younger sample was subsequently supported by Piersel et al. (1977). They found that modeling was superior both to standard administration and to feedback of results on test performance with minority schoolchildren. Failure to establish initial equivalence of treatment groups, however, vitiated results.

Raskind and Nagle (1980), on the other hand, had a more robust design than their predecessors. The lack of significant results reported by Raskind and Nagle might have been due to the limited size of the experimental cells, large group variances, and perhaps inappropriate statistical analysis. The most recent investigation (Pollock 1988) which compared verbal praise, a mastery model, and standard test administration was also problematic. The relatively low number of subjects per treatment, brief exposures to the tapes, and use of a mastery model might have adversely affected results.

In short, the dearth of studies and the problems besetting research on the effects of modeling experiences on the cognitive test scores of children render judgments regarding the utility of vicarious experiences premature.

Statement of the Problem

Although various forms of incentives have been used to facilitate performance, the effects of type of examiner comments (i.e., verbal praise and neutral comments) and type of modeling (i.e., modeling videotape with narration, modeling videotape without narration, and no modeling) on children's cognitive test scores have not been explored.

Purpose of Study

The purpose of this investigation was to compare six different modes of WISC-R test administration. In so doing, examiners presented administrative regimes that involved *in vivo* examiner comments and symbolic modeling.

Rationale

Identification of children in need of special education services is based on information that is generally derived from various sources (e.g., parental interview regarding developmental history, classroom observations, teacher consultations, and psychoeducational evaluation). Nevertheless, intelligence test results represent a major factor in decisions concerning classification, placement, and intervention (Reschly, 1981). Eliciting the best rather than the usual performance of a child during formal cognitive testing is a continuing concern for school psychologists for two reasons. First, maximal performance may attest to the greater ability or competence than what the child exhibits in the everyday classroom situation. Second, accurate assessment involves the reduction of systematic error in test scores. The level of motivation has been identified as an important source of error variance during the

administration of individually administered intelligence tests (Scarr, 1981; Zigler & Butterfield, 1968). Thus, attempts to improve motivation via reinforcement techniques and modeling may serve to reduce this form of error.

CHAPTER II

METHODOLOGY

Subjects

Pupils from fourth grade classes in four Roman Catholic elementary schools were invited to participate in the study. The schools were located in predominantly white, middle class neighborhoods in Brooklyn and Queens (City of New York Department of City Planning, 1986). Permission was obtained from the building principals (in consultation with classroom teachers) to use the schools as test sites. The pool of potential subjects consisted of 168 children. Initially, the classroom teachers distributed cover letters and consent forms to the students. (See Appendices A and B for copies of cover letter and consent form.) The latter, which required the signature of both student and parent, had been prepared in accordance with the guidelines of Ethical Principles in the Conduct of Research with Human Participants (American Psychological Association [APA], 1988). The consent form described the following aspects: the purpose of the proposed study, the voluntary nature of participation and withdrawal, a description of risks and benefits, a guarantee of confidentiality, and a debriefing procedure.

A total of 98 pupils or 58% of the combined enrollments consented to take part in the study. Complete sets of protocols on 86 subjects (52%) were eventually obtained. Sampling mortality was due to four factors: the absence of three students; exclusion of one child who had been referred for special education testing; an incomplete protocol of one student; and an unforeseen shortening of the school day at one site where the final examination schedule of the school curtailed testing the last seven subjects. The subjects (39 boys and 47 girls) ranged in age from 9 years, 0 months to 10 years, 6 months. Their mean age was 9 years, 8 months with a standard deviation of 4.76 months. The sample was composed of 81 white, 3 Asian, and 2 Hispanic pupils.

Pre-Test

Otis-Lennon School Ability Test. Initially, the Otis-Lennon School Ability Test (OLSAT), Elementary Level Form R (Otis & Lennon, 1982) was administered to all subjects. According to the test manual, the OLSAT is a standardized test, suitable for group administration, that taps verbal comprehension, as well as verbal, figural, and quantitative reasoning. The OLSAT was used because of its high correlation (r = .85) with the Full Scale WISC-R (Otis & Lennon, 1982) for

grades three through nine. Test time was 45 minutes, plus approximately 15 minutes for the presentation of instructions. Answer sheets were initially scored by the author and were subsequently machine scored prior to data analysis.

Tests Presented as Part of the Modeling Process

Kaufman Assessment Battery for Children. The Kaufman Assessment Battery for Children (K-ABC) (Kaufman & Kaufman, 1983a) is an individually administered intelligence test that was standardized on exceptional and normal children, ages 2 1/2 through 12 1/2 (Kaufman & Kaufman, 1983b).

Selected Subtests of the K-ABC

Gestalt Closure. This subtest requires that the child name the object or activity depicted in a partially completed line drawing. The unique abilities tapped by this measure include perceptual closure, perceptual inference, and conversion of abstract stimuli (Kaufman & Kaufman, 1983b).

Triangles. This measures requires the subject to assemble yellow and blue rubber triangles to match a larger triangle. Kaufman and Kaufman (1983b) indicate that this subtest assesses nonverbal concept formation.

Selected items from the Gestalt Closure and Triangles subtests appeared in a modeling videotape that

was shown to four of the six experimental groups. The decision to use the K-ABC subtests was based upon two considerations. First, the stimulus pictures and the materials of the chosen subtests are "videogenic," i.e., they are large and provide sufficent color contrast so that the viewer can easily observe them. This advantage may enable the viewer to determine item difficulty and to know whether the model has succeeded in the activity. In addition, the Gestalt Closure and Triangles subtests are similar, though not identical, to the Picture Completion and Block Design subtests of the WISC-R (i.e., two of the dependent variables). As such, it was initially assumed that the subjects would perceive that the activities that they were to address would be comparable in terms of difficulty level to the tasks that were presented to the models.

Dependent Measures

Wechsler Intelligence Scale for Children-Revised.

The Wechsler Intelligence Scale for Children (WISC-R) (Wechsler, 1974) is an individually administered battery that consists of five Verbal and five Performance subtests plus two optional subtests. Wechsler (1974) explained the rationale for the diversity of measures as follows: "The WISC-R emphasizes the importance of probing intelligence in as many different ways as

possible, that is, by as many different kinds of tests as one can marshal. It does so also because it assumes ... that intelligence is best regarded not as a single unique trait but as a composite or global entity" (pp. 5-6).

Selected WISC-R Subtests

Arithmetic. This timed test consists of 18 word problems. For children above the age of 8, the starting point is item 5. Problems 1 through 14 are read aloud by the examiner; the remaining problems are read orally by the examinee. According to Glasser and Zimmerman (1967), the test taps the ability to utilize abstract concepts of number as well as the ability to perform numerical operations.

Digit Span. This measure requires that the examinee repeat a series of numbers presented orally by the examiner. The subtest consists of two tasks: Digits Forward which has strings of digits ranging in length from three to nine digits, and Digits Backwards which spans two to eight digits. This test measures short term auditory memory and attention (Glasser & Zimmerman, 1967). Sattler (1988) reports that Digit Span taps two discrete functions: Digits Forward primarily involves rote learning memory whereas Digits Backward requires

considerably greater transformation of the stimulus input prior to recall.

Picture Completion. This subtest consists of 26 line drawings that are presented in a spiral booklet. Children over the age of 8 begin with the fifth item. Examinees must identify the missing element within 20 seconds in order to earn credit. This subtest purports to measure the child's ability to discern between essential and non-essential detail (Glasser & Zimmerman 1967).

Block Design. This timed subtest contains 11 items and necessitates the replication of abstract designs through manipulation of red and white blocks. Examinees who are 8 years or older begin with the third design. In presenting the subtest, examiners model the construction of the initial design. Examinees must subsequently replicate the remaining designs on the basis of pictures that are presented in a spiral booklet. The Block Design subtest is regarded as "a nonverbal concept formation task requiring perceptual organization, spatial visualization, and abstract conceptualization" (Sattler, 1988, p.159).

The WISC-R was selected as the dependent variable because it is the most commonly used individual cognitive measure for children in the United States

(Lubin, Larsen, & Matarazzo, 1984). Thus, results from this investigation may have a degree of external validity beyond the experimental milieu. The decision to use Picture Completion, Arithmetic, Block Design, and Digit Span subtest was based upon several factors. Initially, the use of these measures allowed the investigator to compare her results to the results of related studies (Bergan, McManis, and Melchert, 1971; Bornstein, 1968; Saigh, 1981; Saigh & Payne, 1976, 1979 Witmer, Bornstein, and Dunham, 1971;) that used some or all of the WISC or WISC-R subtests. Moreover, the Arithmetic and Digit Span (and to a lesser extent, Picture Completion) subtests were chosen because they are purported to be especially sensitive to situational variables like anxiety and distractability (Glasser & Zimmerman, 1967; Kaufman, 1979). The Block Design Subtest was selected because it has the highest correlation of the five Performance subtests with the WISC-R Performance (r = .82) and the Full Scale Score (r =.75) (Wechsler, 1974).

Materials

Modeling Videotape

Rehearsed and edited vignettes of two white fourth graders (a girl and a boy) at work on selected items from the Gestalt Closure and the Triangles

subtests of the K-ABC were videotaped. (See Appendix C for transcript of the videotape). The examiner in these scenes was a white, male doctoral student in the Educational Psychology Program of the City University of New York's Graduate School. In brief, the scenes depicted pupils who, despite difficulty with initial items, persevered in the standardized testing situation. The school psychologist, though supportive, made no comments regarding the performance of the pupils. Total viewing time was approximately ten minutes.

Onscreen Narration for Modeling Videotape

An on-camera narrator (a white female, master's level secondary school teacher) provided commentary. (See Appendix D for the actual transcript.) The commentator's remarks were inserted at three points during the videotape (i.e., prior to the two modeling scenes and at the conclusion of the videotape). The narrator highlighted the test behavior of the models (e.g., the models gave incorrect responses to the first items, but they later paid attention and concentrated on the task at hand). The purpose of the narration was to underscore salient model characteristics (e.g., age and behavior) with which subjects could identify. Totally viewing time was approximately three minutes.

Equipment

Each examiner used a 19 inch color televison and a videocassette recorder, provided by the school or the examiners, during the experiment. Digital stopwatches (Westclox model 37431 and Armitron Sports Timer) were used to time WISC-R subtests.

Examiners

A white male and a white female who are enrolled in the Educational Psychology Doctoral Program of the City University of New York's Graduate Center served as examiners. These individuals were certified school psychologists who by training and experience were familiar with the WISC-R administration and scoring.

Procedure

After the investigator received the signed consent forms, The OLSAT was administered to the students in a group testing session by one of three female proctors: two doctoral students in the Educational Psychology Program of the City University of New York's Graduate Center, and an elementary school teacher. The OLSAT School Ability Index (SAI) scores ranged between 82 and 147. The mean score was 110.18 with a standard deviation of 12.83. According to the test manual, the observed mean score lay at the upper end of the "average" range. At this point, subjects were matched

by sex, OLSAT scores, and schools and randomly assigned to one of six treatment conditions. Further statistical analysis revealed no significant OLSAT differences between schools or experimental groups. Thus, the six experimental groups were matched with respect to SAI prior to treatment.

Treatments

One day to two weeks after the completion of the OLSAT testing, the second phase of the study was implemented. Examiners conducted the experiment in separate rooms. No other activity took place in the room while the experiment was being conducted. All subjects were introduced to the experimental condition by the examiner in the following manner:

"Hello, (Name of student). My name is (Name of examiner). I'm going to give you some tests to see what you're good at and what you need more help with." At this point the following treatments were presented.

I. Modeling Videotape with Onscreen Narration Followed by WISC-R Administration with Examiner Verbal Praise

The examiner continued by stating, "But before we get started, we're going to watch a videotape." Subjects viewed the modeling videotape which contained the onscreen narration. Pursuant to the videotape presentation, the WISC-R Arithmetic, Picture Completion,

Block Design, and Digit Span subtests were administered. During the administration, the examiner issued the following statements after the first three responses in each subtest: "That's the way to try," "That's the way to concentrate," and "Good try; keep on working." Every third response after completion of the first three items, the examiner commented, "See how focusing helps you with the task," "I can see that you're putting a lot of effort into it," or "You're really working hard." Between subtests, the examiner stated, "Paying attention has really helped. Keep up the good work on the next part." The response was emitted regardless of the correctness of the subject's response or solution.

II. Modeling Videotape without Narration Followed by WISC-R Administration with Examiner Verbal Praise

The examiner continued by stating, "But before we get started, we're going to watch a videotape." Subjects in this treatment group viewed the modeling videotape, but the videotape did not include the onscreen narration. After the videotape, the WISC-R subtests were presented and the examinee's responses were verbally praised as in Treatment I.

III. No Modeling and WISC-R Administration with Examiner Verbal Praise

Subjects in this treatment group did not view a modeling videotape. The WISC-R subtests were immediately presented and the examinee's responses were reinforced with the same examiner verbal praise statements as in Treatment I.

IV. Modeling Videotape with Onscreen Narration Followed by WISC-R Administration with Examiner Neutral Comments

The examiner continued by stating, "But before we get started, we're going to watch a videotape." Subjects in this treatment group viewed the same videotape with narration as the children in Treatment I. During administration of the WISC-R subtests the examiner stated, "Now try this," "How about this," and "Here is the next one," after each of the first three responses in a subtest. Every third response after completion of the first three items, the examiner commented, "OK," "Give this one a try," and "Let's try this." Between subtests, the examiner remarked, "Let's try something different." The response was emitted regardless of the correctness of the subject's response or solution. In effect, the neutral comments controlled for verbal praise.

V. Modeling Videotape without Narration Followed by WISC-R Administration with Examiner Neutral Comments

The examiner continued by stating, "Before we get started, we're going to watch a videotape." Subjects in this treatment group viewed the same modeling videotape as their counterparts in Treatment II. After this, the WISC-R subtests were presented and the examinee's responses were reinforced with the same examiner neutral comments as those given to subjects in Treatment IV.

VI. No Modeling and WISC-R Administration with Examiner Neutral Comments

Subjects in this treatment group did not view a modeling videotape. The WISC-R subtests were immediately presented and the examinee's response was reinforced with the same examiner neutral comments statements as in Treatment IV.

The sequence of administration of the four subtests was counterbalanced for all treatment groups in order to control for subtest presentation. Responses and scores were entered on a specially designed answer sheet in which the subtests were arranged in the same order as test administration.

Examined conceptually, Figure 1 presents a schematic representation of the data collection design.

Figure 1. Data collection design.

Treatments	WISC-R Subtests: Arithmetic	Picture Completion	Block Design	Digit Span
I. Modeling videotape with narration and WISC-R examiner verbal praise (n=14)				
II. Modeling videotape without narration and WISC-R examiner verbal praise (n=15)				
III. No modeling and WISC-R examiner verbal praise (n=15)				
IV. Modeling videotape with narration and WISC-R examiner neutral comments (n=15)				
V. Modeling videotape without narration and WISC-R examiner neutral examiner comments (n=14)				
VI. No modeling videotape and WISC-R examiner neutral comments (n=13)				

Research Design and Hypotheses

The design for this study was a 2 x 3 arrangement with the following independent variables: WISC-R examiner comments (i.e., verbal praise and neutral comments) and modeling (i.e., videotaped modeling with narration; videotaped modeling without narration; and no modeling). The outcome variables were the WISC- R Arithmetic, Picture Completion, Block Design, and Digit Span scaled scores.

The following directional hypotheses were tested:

H_{01}: There would be a main effect for verbal praise, i.e., the mean scaled scores of subjects who were verbally praised would be significantly higher than the mean scaled scores of subjects who were not verbally praised across the four WISC-R subtests.

H_{02}: The mean scaled scores of subjects who viewed modeling videotapes would significantly exceed the mean scaled scores of subjects who did not view a modeling videotape.

H_{03}: The mean scaled scores of subjects who viewed the modeling videotape with narration would significantly exceed the mean scaled scores of subjects who viewed the modeling videotape without narration.

In addition to Hypotheses 1, 2, and 3, the interaction effect between the two levels of examiner

comments (i.e., praise and no praise) and the three levels of modeling (i.e., modeling videotape with narration, modeling videotape without narration, and no modeling videotape) was tested.

CHAPTER III

RESULTS

The WISC-R administrations for Treatments I through VI were audiotaped. Half of the audiotapes for each examiner were randomly selected and rated by two independent raters (doctoral students in the Educational Psychology Program of the City University of New York's Graduate School and in the Community Psychology Program of New York University) for compliance with the stated protocol. Viewed along these lines, each rater checked the appropriateness of the examiner's remarks or silence after each of the subjects' responses on the Arithmetic, Picture Completion, and Digit Span subtests. Each rater also determined if the examiners' remarks complied with the experimental protocol after the first three solutions to the Block Design subtest. Interrater reliability estimates were calculated on the basis of the number of agreements divided by the total number of disagreements and disagreements times 100. Examined along these lines, interrater agreement was estimated at 99%.

The WISC-R protocols were scored by an independent doctoral level clinical psychologist who was not aware of the research hypotheses. A second independent examiner (a doctoral student in the Educational

Psychology Program at the City University of New York's Graduate School) checked the scored protocols as well. Interrater agreement for scoring compliance was estimated at 95%. In those instances where discordant decisions were made, the principal investigator individually reviewed the test protocol and decided on the test score.

Tables 1 through 4 present a breakdown of WISC-R subtest scores for each of the six treatment groups wherein it is apparent that all of the mean scaled scores were well within the average range as based on the WISC-R normative data.

Table 1
Mean Scaled Scores and Standard Deviations on the WISC-R Arithmetic Subtest

	Modeling Conditions		
	Modeling + Narration	Modeling Only	No Modeling
Examiner Comments			
Verbal Praise	M 11.89 SD 2.38 (n=14)	M 11.67 SD 2.38 (n=15)	M 11.60 SD 2.77 (n=15)
Neutral Comments	M 12.33 SD 3.06 (n=15)	M 13.15 SD 3.08 (n=14)	M 11.33 SD 3.11 (n=13)

Table 2
Mean Scaled Scores and Standard Deviations of the WISC-R Picture Completion Subtest

	Modeling Conditions		
	Modeling + Narration	Modeling Only	No Modeling
Examiner Comments			
Verbal Praise	M 11.93 SD 2.06 (n=14)	M 10.87 SD 2.29 (n=15)	M 12.40 SD 2.50 (n=15)
Neutral Comments	M 12.37 SD 1.87 (n=15)	M 11.54 SD 2.63 (n=14)	M 11.53 SD 1.92 (n=13)

Table 3
Mean Scaled Scores and Standard Deviations of the WISC-R Block Design Subtest

	Modeling Conditions		
	Modeling + Narration	Modeling Only	No Modeling
Examiner Comments			
Verbal Praise	M 11.00 SD 4.62 (n=14)	M 10.33 SD 1.99 (n=15)	M 11.47 SD 2.26 (n=15)
Neutral Comments	M 10.73 SD 2.79 (n=15)	M 10.85 SD 3.74 (n=14)	M 10.33 SD 3.72 (n=13)

Table 4
Mean Scaled Scores and Standard Deviations of the WISC-R Digit Span Subtest

	Modeling Conditions		
	Modeling + Narration	Modeling Only	No Modeling
Examiner Comments			
Verbal Praise	M 11.00 SD 2.15 (n=14)	M 10.67 SD 2.58 (n=15)	M 11.67 SD 4.20 (n=15)
Neutral Comments	M 10.67 SD 7.74 (n=15)	M 12.77 SD 2.69 (n=14)	M 10.27 SD 3.88 (n=13)

A 2 (praise vs. neutral comments) X 3 (modeling with narration, modeling without narration, and no narration) multivariate analysis of covariance (MANCOVA) was computed on the WISC-R scaled scores with the OLSAT scores as covariates. MANCOVA combines the multivariate analysis of variance (MANOVA) procedure (whose basic structure is similar to its univariate counterpart) (Haase & Ellis, 1986) with the analysis of covariance (ANCOVA). Use of a multivariate procedure enables the researcher to analyze all dependent variables simultaneously (Spector, 1981). The multivariate technique also has two principal advantages. First, it maintains an alpha level not affected by the number of dependent variables. Second, the procedure avoids artificial inflation of group differences due to intercorrelations among the dependent variables (Yaremko, 1982). The value of the covariance technique lies in the partitioning of variation into three, instead of two, components (i.e., variation explained by the treatments, variation explained by the covariate, and residual or unexplained variation [Porter & Raudenbush, 1987]). Table 5 presents the MANCOVA results.

Table 5
Multivariate Analysis of Covariance for Four WISC-R Subtests

Source	Wilk's Lambda	F	Significance
Examiner Comments	.97	.53	.712
Comments x Modeling	.91	.91	.513
Modeling vs. No Modeling	.92	1.65	.171
Narration vs. No Narration	.94	1.24	.302

The main effect for praise, stated in H_{01} was examined and no significant differences were observed, F (4, 76) = 0.53, p>.05. H_{02} stated that the mean scaled scores of the subjects who viewed the modeling videotapes would exceed the mean scaled scores of the subjects who did not view modeling videotapes. Data analysis failed to support this hypothesis, F (4, 76) = 1.71, p>.05. H_{03} stated that the mean scaled scores of the subjects who viewed the modeling videotape with narration would significantly exceed the mean scaled scores of the subjects who viewed the modeling videotape without narration. Data analysis failed to denote a significant difference for this level of comparison, F (4,76) = 1.24, p>.05. In a similar vein, no significant interaction effects were apparent, F (8, 152) = 0.91, p>.05. In view of the non-significant differences, follow-up tests were not conducted.

CHAPTER IV

SUMMARY AND CONCLUSIONS

Summary

The present study compared the effects of two types of examiner comments (i.e., non-contingent verbal praise and netural, non-evaluative comments) and three types of modeling experiences (i.e., videotaped modeling of test behavior, accompanied by onscreen narration; videotaped modeling of test behavior without narration; and no modeling) on the selected WISC-R subtest performance of fourth grade students. The subject pool consisted of 86 primarily white (97%) fourth grade pupils attending urban Roman Catholic parochial schools. Subjects were matched by sex and OLSAT scores and randomly assigned to one of six experimental treatments. Data analysis failed to evince significant effects for praise, modeling, modeling with narration, or interaction. A general discussion regarding the observed results, limitations of the study, and directions for future research will be presented in this chapter.

Discussion

Verbal Praise

The first hypothesis predicted a main effect for examiner praise, i.e., that the mean scaled scores of subjects who received noncontingent verbal praise would

significantly exceed the scaled scores of the subjects who received neutral comments. Statistical analysis failed to support the prediction. Possible explanations for the lack of significant results vis-à-vis earlier studies (Bornstein, 1968; Mayfield, 1977; Saigh, 1981; Saigh & Payne, 1976, 1979; Witmer et al., 1971) may have to do with the actual examiner statements and the characteristics of the selected sample.

Although the verbal praise statements (e.g., "That's the way to try," "That's the way to concentrate," "Good try; keep on working," "See how focusing helps you with the task," "I can see that you're putting a lot of effort into it," and "You're really working hard") were presented in an effort to direct attention to task relevant behavior, it may be argued that these statements were poor motivators. It should be recalled in this context that the studies that observed significant WISC-R effects for praise (Bornstein, 1986; Mayfield, 1977; Saigh, 1981; Saigh & Payne, 1976, 1979) used more emphatic statements (e.g., "Good," "Fine," "That was fine," "Very good; keep up the effort"). As such, the more positively valenced statements might have induced a greater degree of effort than the statements that were presented herein.

Examined from a different perspective, it may also be said that the subject's attributions for success might have affected the results. In his review of teacher praise, Brophy (1981) observed that the frequency for verbal praise is relatively low in the classroom. Brophy also noted that teachers tend to praise low achieving students (for effort rather than achievement) more frequently than high achieving students. Viewed along these lines, Pearl (1985) observed that if "childen are repeatedly told they succeed because they tried hard, they may come to infer that they must be lacking in ability (p.447)." With these points in mind, the experimental subjects might have perceived the examiner verbal praise statements as a form of negative feedback and these perceptions might have served to limit performance.

The non-significant effects for praise may also be related to the characteristics of the selected sample. It should be recalled in this context that the selected sample consisted of non-referred, primarily white middle class fourth grade parochial school students with a mean Otis-Lennon IQ of 110. Moreover, the selected subjects were not experiencing academic or behavioral problems. In addition, the school administrators and teachers reported that the subjects were accustomed to working at

optimal level during examinations (an observation shared by the principal investigator). On the other hand, the subjects in Mayfield's (1977) study differed from the current sample on a number of important variables. More specifically, Mayfield's cases were behaviorally disordered children who were receiving mental health services at a psycho-diagnostic center. Moreover, their mean IQ was 98. In a similar vein, the subjects in the WISC-R studies that were conducted by Saigh (Saigh, 1981; Saigh & Payne, 1976, 1979) also differed from the subjects who were studied herein. All of the subjects in Saigh's studies were educably mentally retarded children who were institutionalized in state facilities. With these points in mind, it may be argued that the examiner verbal praise procedure had more of an effect on the special education subjects because they were not accustomed to being praised for their work on standardized tests. In contrast, the verbal praise procedures may have been less effective for the parochial school students inasmuch as they were accustomed to performing at an average or an above average level and as they were used to doing so without the benefit of external incentives.

Modeling Effects

The second hypothesis predicted that subjects in the modeling condition would achieve significantly higher scores than subjects who did not view a modeling videotape. Moreover, H_{03} predicted that subjects who viewed the modeling videotape with narration would achieve significantly higher scores than subjects who viewed a modeling videotape without narration. Neither hypothesis was supported by the analysis. The non-significant effects may in part be related to the duration of the treatments and the age of the selected sample.

Viewed along these lines, it may be argued that the subjects may have been too young to learn the target behaviors that were presented during the relatively brief intervention. It should be noted in this context that the modeling as well as the modeling with narration regimens lasted for 11 and 13 minutes. Moreover, each regimen presented a number of test taking strategies, (e.g., attention to the task and perseverance with more difficult test items). It is interesting to note in this context that the previous studies (Pollock, 1988; Raskind & Nagle, 1980) that presented similar modeling regimens to middle class groups of children whose ages and IQs approximated the selected sample's were

associated with non-significant effects. On the other hand, Jaffe and Carlson (1972) reported a significant effect when a sample of university undergraduates observed a 60 minute film that depicted administrations of selected WISC and WAIS subtests.

Limitations

The external validity of the extant results are limited to samples with similar demographic and personal characteristics. It should be recalled in this context that the subjects were predominantly white, middle class, parochial school students with a mean age of 9 years, 10 months. Moverover, these cases had a mean Otis-Lennon SAI of 110 and none of them were special education students. The results should also be tempered with the realization that a short form of the WISC-R was used. As such, it would be tenuous to generalize these results to the Full Scale WISC-R.

Directions for Future Research

With regard to the effects of verbal praise during intellectual assessment, results of the present study are consistent with Brophy's observation (1981) that "The fact that praise can function as a reinforcer does not necessarily mean that it always or even usually does (p.7)." Verbal praise for effort does not appear to be efficacious with academically successful students.

Future investigators may wish to direct their efforts to comparing statements such as, "Good," "Fine," (which may imply successful achievement) with effort statements such as "I like the way you're trying," (which may convey anything from effort's responsibity for success to a consolation prize for failure.) In addition, other types of examiner comments invite exploration. For example, Buckhalt (1987) examined audiotapes of actual WISC-R administrations by school psychologists. He observed that some examiner comments could be construed as praise or simple encouragement without an attributional implication. Other comments, he discerned, fell into five categories with an attibutional basis: task difficulty ("They're getting harder, huh?); effort ("I like the way you kept on trying on that one"); ability or knowledge ("You're good at these, aren't you?"); practice or experience ("You probably haven't studied that yet in school"); and fatigue "You're probably just getting tired"). It behooves researchers to ascertain whether such real life practices of clinicians significantly impact on cognitive test performance. Buckhalt also noted that 40% of all examiner comments occurred during the Picture Completion and Block Design Subtests. As such, researchers may wish to determine whether manipulating

ratios of examiner comments on different subtests will produce statistically significant outcomes.

Future research should also be directed toward the assessment of non-cognitive factors (i.e., effort). Viewed along these lines, it is proposed that such outcome measures as the number of items attempted or duration of responses be included in follow-up studies. In addition, it would be of interest to add a third independent variable (i.e., tangible reinforcement). More specifically, it would be of interest to compare the effects of praise, modeling, and material reinforcement (as based on the subject's reinforcement preference).

The investigator asssumed that the subjects who viewed the modeling videotapes would identify with the models inasmuch as the models were matched with the characteristics of the viewers (i.e., age, sex, race, and parochial school attire). The investigator did not, however, empirically determine if the selected subjects identified with the models. With these points in mind, investigators who pursue this line of inquiry in the future may wish to determine if their subjects actually identified with the models inasmuch as identification is an integral part of the modeling process (c.f., Akamatsu & Thelen, 1974; Bandura, 1977, 1986; Hartup & Lougee,

1975; Rosenthal & Zimmerman, 1978; Schunk, 1987). A technique such as the one that was employed by Schunk, Hanson, and Cox (1987) would be of utility in this regard. Briefly, Schunk et al.'s subjects were asked to rate their degree of perceived similarity to videotaped models before the dependent variable (i.e., an arithmetic test) was presented. Data analysis evinced a significant correlation between the subjects' perceived similarity to the model and subsequent learning outcomes.

Inasmuch as the present sample was atypical of the population who undergo cognitive ability testing, future researchers should consider replicating the study with clinical samples. Such studies should include children who represent the majority of special education population: the learning disabled, speech impaired, educably mentally retarded, and the emotionally impaired (e.g., children with attention deficit-hyperactivity disorder).

APPENDIX A: COVER LETTER TO PARENTS

217 Berkeley Place
Brooklyn, New York 11217
Date

Dear Parent or Guardian,

I am a doctoral student at the City University of New York's Graduate School and University Center. In order to earn my degree I must carry out a research project. I have chosen "Incentive Conditions and the Selected WISC-R Subtest Performance of Elementary School Children" as my topic. This study will be conducted under the supervision of Professor Phillip Saigh.

(Name of principal), principal of (Name of School), and the fourth grade teachers, have graciously agreed to allow the school to be a study site. The attached information sheet outlines the purpose and procedures of the study. As noted in the summary, the project involves minimal disruption to the school routine.

If you are willing to have your child participate, would you and your child please sign the attached consent form and have your child return it to the teacher.

If you decide not to have your child participate, he (she) will not be penalized in any way.

If you have any questions, please feel free to call me at (718) 789-0555 after 6 p.m.

Thank you for considering my request.

Very truly yours,

Mary-Georgia Pollock

APPENDIX B: CONSENT FORM

INFORMATION
"Incentive Conditions and the Selected WISC-R Performance of Elementary School Children"

INVESTIGATOR

Mary-Georgia A. Pollock
Doctoral Student
Educational Psychology Department
The Graduate School and University Center
of the City University of New York
33 West 42 Street
New York, New York 10036-8099

PURPOSE

To investigate various techniques to improve students' performance during standardized testing.

PROCEDURES

The project will consist of two procedures, to be conducted on separate days:

1. A group administered standardized test that will take approximately 45 minutes to complete;
2. The presentation of a 10 minute videotape of students who are taking a test. After the viewing, the students will sit for selected parts of an individually administered standardized test. This procedure will last approximately 30 minutes. Neither session will result in missing significant school work. After all students have been tested, the examiners will visit the classroom to answer questions.

RISKS

Students will not be exposed to any risks. No testing will occur without the consent of the parents/guardians and their children. Any child who agrees to be tested and subsequently decides against doing so may withdraw from the study without any penalty. All test scores will be kept confidential and will not become part of school records. Furthermore, after each test administration, Ms. Pollock will transform scores to data coding sheets that will not

include the students' names. All test data will be stored in a locked file by Ms. Pollock.

BENEFITS

Students who participate may learn new ways to improve their performance in standardized and classroom tests. Pupils who took part in an earlier version of this project appeared to enjoy the individual attention and novel test materials.

QUESTIONS

Should any questions arise before, during, or after participation in the study, families are encouraged to contact Ms. Pollock at (718) 789-0555 after 6 p.m.

CONSENT

I have read the above description of the study and discussed the procedures with my child. I give consent for my son/daughter to participate in the study. In addition, my son/daughter voluntarily consents to participate.

_____ Date _____
(Signature of Parent or Guardian)

_____ Date _____
(Signature of Student)

Student's Date of Birth _____

APPENDIX C: TRANSCRIPT FOR MODELING VIDEOTAPE

Cast

Narrator (N)

Mr. Winston (MW), school psychologist

Chris (C), fourth grade boy

Ruth (R), fourth grade girl

Scene I

(Narration for Scene I from Appendix D to be inserted.)

MW: (Solicitous) Everything OK, Chris?

C: Yeah, well, now. It's just that I don't like to take tests. I do all right in school and on my homework, but I, I (nods head no) just don't like to take tests. (Fidgets in seat.)

MW: Ah uh... (Sympathetically) Many students I know feel the same way. But (encouragingly) just try to do your best.

C: (Weak smile) OK.

MW: (Smiles) OK. We'll get started.

MW: I'm going to show you some pictures, and I want you to tell me what each picture is.

C: OK.

MW: (Flips to Item 6 - chair) What is this?

C: (Shrugs) I don't know.

MW: Want to guess?

C: (Shakes head no.)

MW: Now try this one (Flips to Item 15 - airplane).

C: (Hesitantly) A rowboat?

MW: (Flips to Item 21 - the numeral 5)

C: (Sigh of relief). That's a 5.

MW: (Flips to Item 10 - battleship)

C: (Slight pause) Hmm. (Smiles) Oh, that's a ship, a battleship.

MW: (Flips to Item 19 - gymnast).

C: (Moves closer toward picture; pause; gestures with arms). Oh, it's a girl doing gymnastics.

MW: (Flips to Item 20 - Chinese junk.)

C: (Pause) That's a Chinese boat, I forget what you call it.

MW: (Flips to Item 22 - guitarist) Now try this one.

C: That's a hard one. (Pause). I'm not sure. (Pause. Unsure) A flower?

MW: (Flips to Item 23) And this one?

C: (Slight pause.) Hmm... A flag flapping. (MW starts to turn page). Oh wait, can I change my answer.

MW: If you wish.

C: It's not a flag. It's a man climbing something. Yeah, that's what it is.

MW: (Flips to Item 24 - teapot).

C: This one's easy. (Confidently) It's a teapot. (Smiles)

MW: Now I have something different for you to try.

C: OK.

MW: (Lays out two blue and two yellow triangles. Flips card to Item 11. Picks up stopwatch.) Try to make a triangle just like the one in this picture. (Starts timing.)

C: Oh, that's easy. (Assembles correctly.)

MW: (Takes triangles apart. Flips to Item 12.) Now put this one together. (Starts timing.)

C: OK. (Assembles carefully and correctly).

MW: (Dismantles design. Lays out two more yellow and two more blue pieces. Flips to Item 13.) Now try this. (Starts timing.)

C: (Pauses. Looks at picture. Assembles correctly, EXCEPT for extreme right yellow triangle. Instead, uses blue triangle. Unaware of error.) There, finished!

MW: (Dismantles design. Adds ninth triangle. Flips to Item 15.) Now try this. (Starts timing.)

C: (Assembles slowly and correctly.)

MW: (Dismantles design. Flips to Item 16.) Now try this one. (Starts timing.)

C: That's easy. (Proceeds to work on it, but has difficulty with the end triangles.) It's harder than I thought. (Eventually realizes that triangles at outside of design are actually composed of two triangles.) Finished. I finally got it. (Excited). Where's the next one.

MW: That's the last of the triangles, Chris. What do you think of the test so far?

C: It's not too bad. I still don't like the thought of taking tests. I was kinda nervous at first. Well, I guess I still am. But the test don't bother me as much now.

MW: That's good to hear. Let's move on to the next test. (Begins to flip card to next test. End of scene.)

Scene II

(Narration for Scene II appears in Appendix D.)

MW: Ready to begin?

R: Yes. (Begins looking around room.)

MW: (Flips card to Item 6 - chair.) What is this?

R: (Somewhat startled.) Oh! Ah, that's Chinese (starts looking around again) writing I guess.

MW: (Flips to Item 15 - airplane). Now try (Ruth interrupts)

R: (Quickly and in a somewhat louder voice than before) That's a shark.

MW: (Flips to Item 21 - the numeral 5)

R: That's a doughnut, no, wait... (Looks closer). No, that's a 5.

MW: (Flips to Item 10 - ship).

R: (Settling in to chair. Looks more closely at picture.) This one's hard. (Almost as if talking to self) Dishes on the table, no. (Fingers picture.) It looks like a ship. Some sort of navy ship.

MW: (Flips to Item 19 - gymnast.)

R: (Slight pause.) That's a girl doing handsprings like a gymnast.

MW: (Flips to Item 20).

R: (Slight pause). That's a boat. Yeah (Shakes head yes). A junk.

MW: (Flips to Item 22 - guitarist.) Now try this one.

R: (Pause.) I don't know, I can't tell. (Attention starts to drift.)

MW: (Reassuringly). Want to try to guess?

R: (Taking a second look. Slight pause.) Oh, that's somebody playing a guitar.

MW: (Flips card to Item 23 - mountain climber) And this one?

R: This one's hard, too. (Slight pause.) It looks like ah, like a, maybe it's a giraffe.

MW: (Flips to Item 24 - teapot.)

R: (Pause.) That's a pot, a teapot.

MW: Now I have something different for you to try. (Lays out two blue and two yellow triangles. Flips card to Item 11. Picks up stopwatch.)

R: (Notices stopwatch. Looks somewhat puzzled.) What's that?

MW: (Matter-of-factly). This is a stopwatch. I'm going to make notes about the time.

MW: Try to make a triangle just like the one in this picture.

R: (Starts assembly while saying) I know how to do this. (Correct assembly.)

MW: (Dismantles design. Flips to Item 12.) Now put this one together. (Starts timing.)

R: (Puts pieces together correctly. When completed, looks as MW.) Finished.

MW: Dismantles design. Lays out two more yellow and two more blue pieces. Flips to Item 13.) Now try this. (Starts timing.)

R: (Takes time. Assembles correctly.)

MW: (Dismantles design. Adds ninth triangle. Flips to Item 15) Now try this. (Starts timing.)

R: (Assembles slowly and correctly.) Finished.

MW: (Dismantles design. Flips to Item 16.) Now try this one. (Starts timing.)

R: (Works on design, but can't assemble correctly. Starts to hold pieces against picture.)

MW: Let's see if you can do this without putting the pieces up against the picture.

R: OK. (continues to work. Incorrect solution. Pauses without indicating to MW that she has finished.)

MW: Are you finished?

R: Yes.

MW: That's the last of the triangles. What do you think about what we've done so far?

R: I liked the puzzles. I know I messed up the last one, but it was a hard one.

MW: Yes, it was a hard one.

R: Well, I figure that even if I got some wrong, that wouldn't mean the end of the world.

MW: You're right. Ready for the next test?

R: I'm ready.

End of Scene II

(Closing narration appears in Appendix D.)

APPENDIX D: ON-SCREEN NARRATION FOR MODELING VIDEOTAPE

(To be inserted prior to the modeling videotape script that appears in Appendix C.)

Scene I

N: Hello. You are going to see a videotape of two fourth grade students who recently took a test with Mr. Winston. The first student, Chris, may be like you. He likes to ride his bike and he claims that his favorite subject is lunch, but it's really computer class. But one thing that Chris doesn't like, and he admits to Mr. Winston, is that he doesn't like to take tests. In fact, he's probably more uncomfortable in this session than usual. He doesn't know Mr. Winston, he's never taken a test like this one, and he's being videotaped.

Chris gets off to a poor start. He misses the first two items which are somewhat easy. Now Mr. Winston doesn't expect him to get everything right. He knows that some students will find some questions harder than others.

Once Chris is used to the test, he is able to pay attention. He continues to try even when the test gets more difficult, and even if some of his answers or solutions are wrong. At the end of the test, Chris still hasn't changed his opinion of tests, but at least he's willing to go on. Let's watch.

Scene II

(To be inserted prior to Scene II script.)

Now we're going to see Ruth with Mr. Winston. Ruth is like many fourth graders. She enjoys playing with her friends after school and she can't wait to go swimming at Jones Beach this summer. And when it comes to the Super Mario Brothers with the Nintendo, she's the best player in her family. Like Chris, Ruth starts off poorly in this test. She's not very attentive at the beginning. She's curious about the room; after all, she's never taken a test before while someone was videotaping her. At the beginning of the test Ruth looks around the room instead of looking at the picture. Ruth blurts out an answer without thinking or looking carefully. She also notices Mr. Winston's stopwatch. Ruth does begin to settle down. The camera is no longer a distractor, and Ruth begins concentrating on the pictures and the puzzles. As a result, she is able to answer most items correctly and to make most designs match the pictures. Her test behavior is still not perfect. Ruth realizes that she has made some errors but is able to accept her less-than-perfect score. Let's watch.

Conclusion

(To be inserted at end of Scene II.)

As you just saw, Chris and Ruth sometimes were not sure what the right answer or solution was, but they tried anyhow and were willing to take a good guess. When this videotape ends, you will begin your test.

REFERENCES

Akamatsu, T. J., & Thelen, M. H. (1974). A review of the literature on observer characteristics and imitation. Developmental Psychology, 10, 38-47.

American Psychological Association. (1988). Ethical principles in the conduct of research with human participants. Washington: American Psychological Association.

Anastasi, A. (1987). Psychological testing (6th ed.). New York: Macmillan.

Bandura, A. (1968). Modeling approaches to the modification of phobic disorders. In R. Porter (Ed.), The role of learning in psychotherapy. London: Churchill.

Bandura, A. (1969). Principles of behavior modification. New York: Holt, Rinehart & Winston.

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1986). Social foundations of thought and action: A social-cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.

Benton, A. (1936). Influence of incentives upon intelligence test scores of school children. Journal of Genetic Psychology, 74, 494-497.

Bergan, A., McManis, D. L., & Melchert, P. A. (1971). Effects of social and token reinforcement on WISC Block Design performance. Perceptual and Motor Skills, 32, 871-880.

Binet, A., & Simon, T. (1983). The development of intelligence in children (S. Kite, Trans.). Salem, NH: Ayer. (Original work published 1916)

Bornstein, A. V. (1968). The effects of examiner approval and disapproval upon the performance of subjects on the performance scale of the Wechsler Intelligence Scale for Children. Dissertation Abstracts International, 28B, 3047. (University Microfilms No.68339)

Brophy, J. (1981). Teacher praise: A functional analysis. Review of Educational Research, 51, 5-32.

Brophy, J. (1983). Conceptualizing student motivation. Educational Psychologist, 18, 200-215.

Brophy, J. (1986). Teacher influences on student achievement. American Psychologist, 41, 1069-1077.

Brophy, J., & Everston, C. (1976). Learning from teaching: A developmental perspective. Boston: Allyn & Bacon.

Bussey, K., & Bandura, A. (1984). Influence of gender constancy and social power on sex-linked modeling. Journal of Personality and Social Psychology, 47, 1292-1302.

City of New York Department of City Planning. (1986). Socioeconomic profile: A portrait of New York City from the 1980 Census. New York: Author.

Cronbach, L. J. (1984). Essentials of psychological testing. New York: Harper & Row.

Dunn, L. M. (1959). Peabody Picture Vocabulary Test. Minneapolis: American Guidance Service.

Feldman, S. E., & Sullivan, D. S. (1971). Factors mediating the effects of enhanced rapport on children's performance. Journal of Consulting and Clinical Psychology, 36, 302.

Fish, J. M. (1988). Reinforcement in testing: Research with children and adolescents. Professional School Psychology, 3, 203-218.

French, J., & Murphy, J. (1985). Intelligence, its assessment and its role in comprehensive evaluation. In J. R. Bergan (Ed.), School psychology in contemporary education: An introduction. Columbus, OH: Charles E. Merrill.

Galdieri, A., Barcikowski, R., & Witmer, J. (1972). The effect of verbal approval upon performance of middle class and lower class third grade children on the WISC. Psychology in the Schools, 9, 404-408.

Glasser, A. J., & Zimmerman, I. L. (1967). Clinical interpretation of the Wechsler Intelligence Scale for Children. New York: Grune and Stratton.

Goh, D. S., & Lund, J. M. (1977). Verbal reinforcement, socio-economic status, and intelligence test performance of preschool children. Perceptual and Motor Skills, 44, 1011-1014.

Graham, G. A. (1971). The effects of material and social incentives on the performance of intelligence test tasks by lower-class and middle-class Negro preschool children (Doctoral dissertation, George Washington University). Dissertation Abstracts International, 31B, 4311.

Haase, R. F., & Ellis, M. V. (1986). Multivariate analysis of variance. Journal of Counseling Psychology, 34, 404-413.

Hartup, W. W., & Lougee, M. D. (1975). Peers as models. School Psychology Digest, 4, 11-21.

Havighurst, R. J. (1970). Minority subcultures and the law of effect. American Psychologist, 25, 313-322.

Hurlock, E. (1924). The value of praise and reproof as incentives for children. Archives of Psychology, 11, 74.

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

Jaffe, P. G., & Carlson, P. M. (1972). Modelling therapy for test anxiety: The role of model affect and consequences. Behavior Research and Therapy, 10, 329-339.

Kaufman, A. S. (1979). Intelligent testing with the WISC-R. New York: John Wiley & Sons.

Kaufman, A. S., & Kaufman, N. L. (1983a). Administration and scoring manual for the Kaufman Assessment Battery for Children. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1983b). Kaufman Assessment Battery for Children: Interpretative manual. Circle Pines, MN: American Guidance Service.

Klorman, P., Hilpert, P. L., Michael, R., LaGana, C., & Sveen, O. B. (1980). Effects of coping and mastery modeling on experienced and inexperienced pedodontic patients' disruptiveness. Behavior Therapy, 11, 86-96.

Klugman, S. F. (1944). The effect of money incentive versus praise upon the reliability and obtained scores of the revised Stanford-Binet. Journal of Genetic Psychology, 30, 255-269.

Kornhaber, R. C., & Schroder, H. E. (1975). Importance of model similarity on extinction of avoidance behavior in children. Journal of Consulting and Clinical Psychology, 43, 601-607.

Lubin, B., Larsen, P. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935-1982. American Psychologist, 39, 451-454.

Mayfield, A. V. (1977). The effects of social reinforcement and prior relationship with the examiner on the intellectual performance of children with behavioral disorders. Dissertation Abstracts International, 37, 5040A. (University Microfilms No. 77-1553)

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

Miller, R. A. (1974). Social milieu and the effects of reinforcement on I.Q. tests (Doctoral dissertation, University of North Carolina at Chapel Hill, 1973). Dissertation Abstracts International, 34, 517-B.

Mitchell, J. V., Jr. (Ed.). (1985). The ninth mental measurements yearbook. Lincoln, NB: University of Nebraska Press.

Otis, A. S., & Lennon, R. T. (1982). Otis-Lennon school ability test: Manual for administering and

interpreting (Elementary Form R). San Antonio, TX: Harcourt Brace Jovanovich.

Pearl, R. (1985). Interventions for increasing motivation. Journal of Abnormal Child Psychology, 13, 443-454.

Perry, D. G., & Bussey, K. (1979). The social learning theory of sex differences: Imitation is alive and well. Journal of Personality and Social Psychology, 37, 1699-1712.

Piersel, W. C., Brody, G. H., & Kratochwill, T. R. (1977). A further examination of motivational influences on disadvantaged minority group children's intelligence test performance. Child Development, 48, 1142-1145.

Pollock, M. G. (1988). A comparison of three methods to improve intelligence test performance. Unpublished manuscript, City University of New York Graduate School and University Center.

Porter, A. C., & Raudenbush, S. W. (1987). Analysis of covariance: Its model and use in psychological research (Special Issue: Quantitative Foundations of Counseling Psychology Research). Journal of Counseling Psychology, 34, 383-392.

Quay, L. C. (1971). Language dialect, reinforcement, and the intelligence test performance of Negro children. Child Development, 42, 5-15.

Quay, L. C. (1975). Reinforcement and Binet performance of disadvantaged children. Journal of Educational Psychology, 67, 132-135.

Raskind, L. T. (1978). The effects of three modeling films on the WISC-R scores of test-anxious fifth-graders (Doctoral dissertation, University of South Carolina, Columbia).

Raskind, L. T., & Nagle, R. N. (1980). Modeling effects on the intelligence test performance of test-anxious children. Psychology in the Schools, 17, 351-355.

Reschly, D. J. (1981). Psychological testing in educational classification and placement. American Psychologist, 36, 1094-1102.

Rosenthal, T. L., & Zimmerman, B. J. (1978). Social learning and cognition. New York: Academic Press.

Saigh, P. A. (1981). The effects of positive examiner comments on the total WISC-R performance of institutionalized EMR students. Journal of School Psychology, 19, 87-91.

Saigh, P. A., & Payne, D. A. (1976). The influence of examiner verbal comments on the WISC performance of EMR students. Journal of School Psychology, 14(14), 343-345.

Saigh, P. A., & Payne, D. A. (1979). The effect of type of reinforcer and reinforcement schedule on performances of EMR students on four selected subtests of the WISC-R. Psychology in the Schools, 16, 106-110.

Sarason, I. G. (1971). Experimental approaches to test anxiety: Attention and uses of information. In C. D. Speilberger (Ed.), Anxiety and behavior (Vol. 2). New York: Academic Press.

Sarason, I. G., Davidson, K. S., Lightall, F. F., Waite, R. R., & Roebush, B. K. (Eds.). (1960). Anxiety in elementary school children. New York: Wiley.

Sattler, J. (1988). Assessment of children (3rd ed.). San Diego, CA: Jerome M. Sattler.

Scarr, S. (1981). Testing for children: Assessment and the many determinants of intellectual competence. American Psychologist, 36, 1159-1166.

Schunk, D. H. (1987). Peer models and children's behavioral change. Review of Educational Research, 57, 149-174.

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self efficacy and achievement. Journal of Educational Psychology, 77, 313-322.

Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. Journal of Educational Psychology, 79, 54-61.

Silverstein, A. B. (Ed.). (1968). Validity of a new approach to the design of WAIS, WISC, and WPPSI short forms. Journal of Consulting and Clinical Psychology, 32(478-479).

Skinner, B. F. (1953). Science and human behavior. New York: Macmillan.

Spector, P. E. (1981). Research designs. Beverly Hills, CA: Sage.

Stallings, J. (1975). Implementation and child effects of teaching practices in Follow Through classrooms. Monographs of the Society for Research inn Child Development, 40(Nos. 7 and 7 (Serial No. 163)).

Terman, L. M., & Merrill, M. A. (1973). Stanford-Binet Intelligence Scale manual for the third revision: Form L-M. New York: Houghton-Mifflin.

Terrell, F., Taylor, J., & Terrell, S. L. (1978). Effects of type of social reinforcement on the intelligence test performance of lower-class black children. Journal of Consulting and Clinical Psychology, 46, 1538-1539.

Terrell, F., Terrell, S. L., & Taylor, J. (1980). Effects of race of examiner and type of reinforcement on the intelligence test performance of lower-class black children. Psychology in the Schools, 17, 270-272.

Terrell, F., Terrell, S. L., & Taylor, J. (1981). Effects of type of reinforcement on the intelligence test performance of retarded black children. Psychology in the Schools, 18, 225-227.

Thelen, M. H., Fry, R. A., Fehrenbach, P. A., & Frautschi, N. M. (1979). Therapeutic videotape and film modeling: A review. Psychological Bulletin, 86, 701-720.

Tiber, N., & Kennedy, W. A. (1964). The effects of incentives on the intelligencce test performance of

different social groups. Journal of Consulting Psychology, 28, 187-189.

Tufano, L. G. (1976). The effect of effort and performance reinforcement on WISC-R scores of black and white EMR boys (Doctoral dissertation, University of Georgia, 1975). (Dissertation Abstractions International, 36, 5961A)

Vernon, D. T. A. (1974). Modeling and birth order in responses to painful stimuli. Journal of Personality and Social Psychology, 29, 794-799.

Wechsler, D. (1955). Manual for the WAIS. New York: Psychological Corp.

Wechsler, D. (1958). The measurement and appraisal of adult intelligence. Baltimore, MD: Williams & Williams.

Wechsler, D. (1967). Manual for the Wechsler Preschool and Primary Scale of Intelligence. New York: Psychological Corporation.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children - Revised Edition. New York: Psychological Corporation.

Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. American Psychologist, 44, 98-104.

Wienges, M. A. E. (1971). Reinforcement, race and socioeconomic level as factors in intelligence test performance (Doctoral dissertation, University of Pittsburgh, 1971). (Dissertation Abstracts International, 1971, 32B, 7299)

Witmer, J. M., Bornstein, A. V., & Dunham, R. M. (1971). 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. Journal of School Psychology, 9, 347-356.

Yaremko, R. M. (Ed.). (1982). Reference handbook of research and statistical methods in psychology: For students and professionals. New York: Harper & Row.

Zigler, E., & Butterfield, E. (1968). Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. Child Development, 39, 1-14.

Zimmerman, B. J., & Rosenthal, T. L. (1974). Observational learning of rule-governed behavior by children. Psychological Bulletin, 81, 29-42.