INFORMATION TO USERS

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

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

THE EFFECTS OF DIFFERENT INCENTIVE CONDITIONS ON THE WISC-III PERFORMANCE OF CONDUCT DISORDER ADOLESCENTS

by

Edmond Fallon

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.

2001

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

08-13-01
Date

Philip A. Saigh
Chair of Examining Committee

8/27/2001
Date

Executive Officer

Philip A. Saigh, Ph.D.

Alan L. Gross, Ph.D.

Georgiana S. Tryon, Ph.D.
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

THE EFFECTS OF DIFFERENT INCENTIVE CONDITIONS ON THE WISC-III PERFORMANCE OF CONDUCT DISORDER ADOLESCENTS

by

Edmond Fallon

Advisor: Professor P. A. Saigh

Sixty-three conduct disorder youths were randomly selected and assigned to one of three treatment conditions (i.e., material reward, verbal praise, and neutral, non-evaluative comments groups) to study the effects of different incentive conditions on the WISC-III performance on adolescents with conduct disorder. The youths in the material reward condition were noncontingently rewarded with tokens that were exchanged for reinforcers for their WISC-III responses. The participants in the other conditions were verbally rewarded or received neutral comments following the same schedule. Two multivariate analysis of variance procedures revealed no significant differences between the mean WISC-III IQ scores and subtest scores respectively for treatment conditions. These results were inconsistent with the clinical literature that supports the hypothesis that adolescents with conduct disorder perform better on an

individually administered intelligence test when given material rewards in comparison to groups who receive verbal praise or neutral comments. Significant differences ($p < .01$) for the mean total scores of both the material reward and verbal praise groups in comparison to the standard group were observed on the Post Hoc Follow-up Questionnaire using one-way ANOVA and Bonferroni post hoc tests. Material rewards and verbal praise were equally effective in fostering participants' positive, subjective thoughts and feelings associated with the testing experience.

Acknowledgements

A doctoral dissertation cannot be completed without the support, assistance, and encouragement of many individuals whom I must acknowledge. Dr. Saigh provided inspiration, guidance, and direction in his role as teacher, department head, and advisor. I am also grateful for the support and advice of my other committee members, Dr. Gross and Dr. Tryon. I wish to thank Dr. Fish and Dr. Li Puma for their assistance in the completion of my dissertation process. Furthermore, I am indebted to the faculty and my fellow students in the Educational Psychology Program at CUNY for making my training meaningful, current, and relevant.

I would like to thank my colleagues who have been supportive through my dissertation process. In particular, I would like to thank Dr. Frank Egan, my supervisor at Leake & Watts School, and Dr. Swan Brown Ronde, my internship supervisor, for their support, advice, and assistance in making this project a reality. I am also grateful for the assistance of coworkers and friends who served as examiners and raters for my dissertation project.

Finally, I must acknowledge and express my deepest gratitude to my family. I thank both of my parents, Ray and Rose Fallon, for the formative experiences that allowed me to persist in the pursuit of this goal. Last, but not least, I found my

greatest source of strength and support in my wife and children. I could not have completed this project without the constant patience and encouragement offered by my wife Joan. My completion of this dissertation project is an endeavor that I share with my family, friends, and colleagues.

Table of Contents

List of Tables

List of Figures

THE EFFECTS OF DIFFERENT INCENTIVE CONDITIONS ON THE WISC-III PERFORMANCE OF CONDUCT DISORDER ADOLESCENTS

INTRODUCTION

Incentive Conditions in Testing

The purpose of psychological testing is to accurately predict a relatively broad and significant area of behavior based on an objective and standardized measure of the behavior observed during testing (Anastasi, 1997). Traditionally, it has been assumed that an examinee's motivation to perform during testing was optimal (Thorndike, 1924). Thorndike acknowledged that in practice, it was difficult to know if an examinee had shown optimal effort. He suggested that it was "theoretically possible to arrange a system of incentives such that each person ... would put forth approximately his maximum effort" (p. 228). Cronbach (1984), Reschly (1981), and Sattler (1992) stressed the need for examinees to present their best effort during standardized testing situations.

A standard approach to establish and maintain motivation during testing suggested by Anastasi (1997) involves urging examinees to show interest in the test, to cooperate with the examiners, to concentrate on tasks, and to put forth their best effort to perform well. The value of praise to enhance rapport and to maintain high levels of motivation and effort was recognized by pioneers in the

field of psychological testing. Binet and Simon (1916) noted that praise "is indispensable to obtain on their part even a small output of effort" (p. 140). In a similar vein, Terman (1916) stated that "nothing contributes more to satisfactory rapport than praise of a child's efforts ... Statements like 'Fine!', 'Splendid!' etc. should be used lavishly" (p. 125). Terman and Merrill (1973) expressed a more moderate position as they advocated the use of nonspecific praise for effort. Analogously, Wechsler (1991) urged examiners to "praise and encourage the child for effort made" (p. 37). He also cautioned not "to give feedback on whether a particular response is right or wrong" (p. 37).

In contrast to the aforementioned standard approach to establishing motivation during testing, social and token reinforcement have been used as incentives to maintain examinee motivation and performance during testing (Fish, 1988). A host of studies have examined the use these incentives. A selected review of the literature on social and token reinforcement used as incentives during testing is presented.

History

Early Studies on Incentives in Testing

Incentives in the form of praise and reproof were introduced as a manipulated variable during cognitive testing with children by Hurlock (1924). Hurlock matched 408 third-, fifth-, and eighth-grade children based on sex, age, race, and scores on the Otis Intelligence Scale Primary Examination, Form A (Otis, 1924) or the National Group Intelligence Test, Scale B, Form 1 (The author and year of publication was not stated.). The participants were randomly assigned to one of three treatment conditions (i.e., praise, reproof, or no feedback). Hurlock's description of praise and reproof was limited to the report that these conditions involved encouraging and discouraging comments preceding the tests. The no feedback condition involved no departure from the standardized test procedure. The gain scores after the administration of Form B or Form 2 were significantly greater for the participants in the praise and reproof conditions in comparison to the participants in the control group that received no feedback. Hurlock (1925) replicated this study with 132 fifth- and 141 eighth-grade students and similar results were observed.

In a similar vein, Benton (1936) presented results involving the effects of incentives on group intelligence test performance. Two groups of 12 male and 13

female seventh- and eighth-grade students completed the Otis Self Administering Test (Otis, 1918) on a pretest and post-test basis. The experimental group received incentives prior to the second testing and the control group received a standard test administration. Benton made the motivational factor for participants "as strong as possible" (p. 495) by including praise, strong encouragement, knowledge of the results on the pretest, and the promise of a prize for improvement on the post-test as incentives. Benton reported no significant differences between the gain scores of the experimental group and the control group.

The first study involving the efficacy of incentives to motivate examinees to perform at their highest level on individually administered cognitive tests was conducted by Klugman (1944). Klugman compared praise and monetary rewards contingent upon correct responses on the 1937 Revised Stanford-Binet, Form L and Form M (Terman & Merrill, 1937) in a pretest and post-test design with a one-week interval between testing. His participants were 72 students in grades two through seven matched by sex, race, and grade. Form L and Form M were administered according to one of the four testing conditions: (a) praise during both administrations; (b) money during both administrations; (c) praise during the first administration and money during the second administration; (d) money

during the first administration and praise during the second administration. A control group was not employed.

The participants earned between 5 and 15 cents in the monetary reward treatment. The actual statements that constituted praise were not reported. No significant differences between the mean group scores or between the group reliability coefficients were observed. Klugman noted that the mean scores and reliability coefficients were somewhat higher for the monetary treatment groups in comparison to the praise treatment groups.

Contemporary Studies on Incentives on Testing

Reviews of the reinforcement-in-testing literature (Fish, 1988; Kennedy & Willcutt, 1964; Sattler, 1992) have yielded mixed findings regarding the efficacy of incentive conditions on test performance. Kennedy and Willcutt considered studies from 1914 through 1964. The studies reviewed group and individually administered tests of cognitive and academic performance. Fish's analysis included 35 studies from 1944 through 1986 that involved only individually administered intelligence tests. In addition to reviewing studies on the effects of verbal praise, Fish reviewed studies in which the term reward referred to candy, toys, tokens, and knowledge of results. Fish limited his analysis to studies involving non-Special Education children and adolescents. Sattler's review of 34

investigations included normal and handicapped children from various ethnic and social-class groups who received incentives (e.g., praise, candy, or money) or feedback on cognitive test performance.

Because of the mixed findings in the literature regarding reinforcement-in-testing, a detailed analysis of pertinent studies involving the effects of incentive conditions on cognitive test performance was conducted. Computer searches using two databases were completed. The Educational Resources Information Center (ERIC) from 1966 through April, 2001 and PsychINFO from 1984 through September, 2000 were examined. References appearing in Fish, Kennedy and Willcutt, Sattler, and Pollock (1990) were checked. The MEDLINE data base was searched from 1966 through April, 2001 to identify pertinent studies related to youth with conduct disorder and incentive conditions. This process led to the identification of 40 studies involving the use of incentive conditions during individual cognitive testing.

Studies Showing No Effects for Incentive Conditions

Thirteen studies did not support the use of incentives or feedback for improving participants' performance on cognitive tests. The study of reinforcement-in-testing was re-introduced 20 years after Klugman's investigation by Tiber and Kennedy (1964) who looked at the effects of different incentive

conditions on the intelligence test results of 480 second and third grade students from three social groups. The basis for determining social class was not specified for the lower-class white, lower-class black, and middle-class white groups. Equal numbers of participants from each group were randomly assigned to the four incentive conditions (verbal praise, verbal reproof, candy reward, and control). Tiber and Kennedy did not provide operational definitions for praise and reproof. It is also important to note that no pretest was used to establish the group equivalence. The 1960 Stanford-Binet Form L-M (Terman & Merrill, 1960) was administered to the participants with the experimental incentives administered at the end of each subtest. No significant difference between the scores of the incentive groups or interactions between the type of incentive and the social groups were identified.

In a similar vein, Cohen (1970) investigated the effect of verbal approval and material reinforcement on the average time for completion of two modified items on the Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949) Block Design subtest. Seventy-two second and fifth grade lower- and middle-class students (IQ range 85 - 115) were randomly assigned to eight groups. The estimated value of the participants' parents' real estate was used to determine the social class. The sample included a small, unspecified number of black

participants and approximately equal numbers of male and female participants. It is important to note that Cohen made no attempt to control groups for race or gender or to equate groups through pretest performance. Cohen's study did not include a control group. Verbal approval involved the statements, "Very, very good. You've got the idea," "Very good! You're doing very well," and "You did that one very fast" (p. 44). The participants in the material reinforcement group were told "If you do it as fast as you can, there will be a piece of candy for you when you finish" (p. 45). Candy was given to the participants after the completion of the Block Design items. The average time scores for completion of the two modified WISC Block Design items showed no significant difference between incentive conditions. In addition, no interactions between incentive condition, grade level, and social class were observed.

Graham (1971) studied the effects of candy reward, verbal praise, and no reward on the cognitive performance of 96 black nursery and kindergarten children from middle and lower class homes. Class designation was based on parental income. Equal numbers of male and female participants were selected from each class and were randomly assigned to one of the three treatment groups. The verbal praise treatment involved the following examiner statements: "That's good," "That's very good," "That's fine," and "Very good" (p. 22). The verbal and

reward incentives were contingent upon correct responses and were administered within the context of an unspecified partial reinforcement schedule. The participants' Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Wechsler, 1967) IQ scores were estimated from performance on the Information, Animal House, Sentence Memory, and Picture Completion subtests. Three male and two female black examiners administered the measures to male and female participants. Pre-experimental Peabody Picture Vocabulary Test (PPVT) (Dunn, 1959) scores were used as covariates. The data analysis revealed no main effects for the type of incentive. Graham speculated that the individualized attention experienced by all participants may have facilitated their optimal motivation allowing for negligible impact on the participants' level of motivation for praise or material rewards.

The influence of motivation and communication on the intellectual performance of black preschool children was investigated by Quay (1971). Quay randomly assigned 100 three- and four-year-old Head Start students (55 boys and 45 girls) to one of four treatment groups. No control group was established. The groups received one of two incentive conditions (verbal praise or candy) and one of two types of language conditions (standard English or black English) during the administration of the Stanford-Binet Form L-M. Two black male examiners

administered the instrument. The participants who received the candy were unaware that the rewards were received on a contingency basis for correct responses. The participants who received verbal praise were told "That's good" (p. 9).

No main effects for the incentive conditions or the type of test language were reported. Moreover, no significant interaction effects were reported. The absence of a pretest to equate the groups on intellectual functioning may have affected the results. Quay noted that the IQ scores for her participants were close to the normative mean and greater than the scores of black children who had not attended compensatory educational programs. Quay suggested that the participants in her study may have been functioning close to their intellectual limit. She also argued that the incentive or language conditions may have contributed little to their near optimal performance.

Galdieri, Barcikowski, and Witmer (1972) compared the performance of 72 rural middle- and lower-class white third grade students on the WISC. The participants' social class was determined based on the parent information contained in the student's records using the Two-Factor Index of Social Position (Hollingshead, 1957). The purpose of the study was to determine whether a

difference existed between the verbal-approval and no-approval conditions during the test administration.

The verbal approval condition involved three examiner statements. After the first examinee response on each subtest, the examiner stated "Good!" (p. 405) regardless of the correctness of the response. After the first incorrect response on each subtest, the examiner commented, "That was hard, wasn't it? But you are doing good" (p. 405). Between each subtest, a comment was made to convey approval (e.g., "You are doing fine!" [p. 406]) in addition to a transitory statement (e.g., "Let's go on to something else" or "Now let's try these" [p.406]). No comments to convey approval were given to the participants in the neutral group.

Galdieri et al. found no significant difference between the scores of the treatment groups and no significant interaction effect involving treatment and socio-economic status. It should be noted that although equal middle- and low-class youths were randomly chosen for the study and randomly assigned to the treatment conditions, Galdieri et al. did not determine if the groups were equivalent with regard to intellectual functioning before the treatments were presented.

The effects of combined monetary and verbal reinforcement on the speed of writing, copying, and coding tasks were examined by Lyle and Johnson (1973)

through three experiments. Participants were two groups of 40 children (age range 6.75 years - 9.0 years) who had been identified as good coders or poor coders based on their performance on the WISC Coding subtest. All participants had IQ scores within the average range. The participants in the speed of writing and copying experiments were subdivided into younger (6.75 to 7.5 years) and older (8.0 to 9.0 years) age groups and were randomly assigned to reinforcement or non-reinforcement conditions. Forty of the 80 participants were sampled for the speed of coding experiment and were randomly assigned to the reinforcement or non-reinforcement conditions. All groups were reported to be approximately equivalent with regard to IQ, age, and coding performance variables.

In the reinforcement condition, the examiner conspicuously dropped a penny into a bowl and simultaneously stated "Good, keep going" (p. 211) as participants completed a predetermined number of responses. In the three experiments, the participants were instructed either to write as many Xs as possible on gridded paper, to copy shapes in squares located directly below the shapes, or to pair symbols and numbers in the format of the WISC Coding subtest with changed symbol-number pairings. Two-minute time limits were imposed on all performances. The participants' scores for writing Xs served as covariates for their copying task scores. Monetary plus verbal reinforcement resulted in

significantly improved performance on the speed of writing task. A positive effect for reinforcement was not observed on the more complex speed of copying or coding tasks. No interaction effects were identified for reinforcement conditions.

Cook (1973) investigated the effects of verbal feedback and monetary reward conditions on the WISC scores of Hispanic fifth-grade students. Ninety participants received either knowledge of their results or a penny following correct responses. The monetary reward and the verbal feedback conditions did not improve the participants' WISC performance.

Quay (1975) randomly assigned 46 male and 46 female black fourth-grade students (age range 8 years, 11 months - 10 years, 2 months) to two incentive conditions. The participants were residents of an urban, high-poverty area. Quay's study did not include a control group. The participants received an administration of the Stanford-Binet Form L-M by a black male examiner. In the verbal praise condition, the participants received praise in the manner specified in the test manual when an item was passed (e.g., "That's good"). The participants in the material reward condition 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" (p. 133). No significant IQ score differences or interactions were reported.

Clingman and Fowler (1975) investigated the effects of candy reinforcement on the Stanford-Binet Intelligence Test scores in a study involving 36 first- and second-grade students of above average intelligence (IQ range 93 - 134). Participants were randomly assigned to one of three incentive groups and were pretested on Form L. Form M was administered under these experimental conditions. The participants were given: candy contingent on correct responses, candy non-contingently, or not given candy. The participants in the contingent group were given a piece of chocolate candy immediately following each correct response. The participants in a yoked-comparison group received the same number of chocolates before responding to a question regardless of the correctness of the answer. The control participants received no candy. Clingman and Fowler found no statistically significant differences between Form L and Form M test scores for the three treatment groups.

Tufano (1976) matched 30 black and 30 white educably mentally retarded male resource room students (age range 10 years - 14 years) on the basis of previous Wechsler Intelligence Scale for Children - Revised (WISC-R) (Wechsler, 1974) IQ scores. The participants were randomly assigned to one of

two treatment groups: verbal praise contingent on effort and verbal praise contingent on performance during the administration of the WISC-R. The study lacked a control group. Effort was operationally defined by any one of the following behaviors: eye contact with the examiner, verbal attempt to answer the question, attempt to perform the task manually, or leaning forward. The following statements were used by the examiner to reinforce effort: "You are trying hard," "I like your helping me," "I'm so glad you enjoy learning". These comments were offered after the third demonstration of 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 you doing a good job," "I'm glad you did a good job".

Tufano reported no main effects for the treatment on the WISC-R Verbal, Performance, and Full Scale IQs, subtest scores, and factor ability scores. The only differential effect for race and treatment was noted on the Information subtest. Black participants in the effort group performed better than the white participants. Analogously, the white participants in the performance group performed better than the black participants.

The effects of tangible reinforcement on the intelligence test performance of 72 trainable mentally retarded students (age range 9 years 4 months - 20 years, 9 months) was examined by Busch and Osborne (1976). The participants (IQ range 30 - 60) were randomly selected from a special school to include an equal number of males and females. The participants were randomly assigned to either an experimental or a control group. Both groups received administrations of the WISC Arithmetic, Picture Arrangement, and Comprehension subtests and the Lorge-Thorndike Intelligence Test Vocabulary subtest (Lorge & Thorndike, 1957). The participants in the experimental group were given a single piece of candy following each response (regardless of the correctness). The participants in the control group received an administration of the tests in the standard manner that is specified in the test manual. The order of the administration of the tests for each participant was randomized to control for sequence effects. Two male and two female graduate student examiners were randomly assigned an equal number of experimental and control participants and were blind to the previous IQ scores of the participants and the hypothesis of the study.

Busch and Osborne reported that significant differences were found between the reinforced administration and the standard administration on the WISC Picture Arrangement and Arithmatic subtests and the Lorge-Thorndike Vocabulary

subtest. No difference was observed on the WISC Comprenhension subtest. Busch and Osborne concluded that reinforcement techniques facilitated performance. They also noted that the magnitude of the effect of the treatment was relatively slight and may not have been clinically meaningful. They suggested that the relative efficiency of reinforcement during test administration may increase over an extended test session.

The effects of verbal feedback, noncontingent praise, and standard test administration on the intellectual performance of 90 preschool children were examined by Goh and Lund (1977). Using the Revised Occupational Scale from Warner, Meeker, and Sell's Index of Status Characteristics (Miller, 1970), equal size middle- and lower-socioeconomic status groups were established. This included approximately equal numbers of male and female participants. The racial composition of the groups was not identified. The participants were randomly assigned to one of three groups and received an administration of the WPPSI Verbal Scale. One treatment group received a standard administration with no reinforcement. The second treatment group received noncontingent examiner praise involving the statements, "Good," "Very good," or "You're pretty smart" (p. 1012). The third treatment group received verbally reinforced responses from the examiner regarding the correctness of the examinee's

response. Verbal reinforcement involved the following statements: "Correct," "Right," or "That's a good answer" (p. 1012). The schedule of reinforcement was not specified. The response, "don't know" or obviously irrelevant responses were not reinforced. The participants' scores from the PPVT served as covariates for the Wechsler scores.

Goh and Lund reported no significant main effect for the type of verbal reinforcement and no interaction effect between the type of treatment and the participants' social status. They suggested that excessive verbal feedback may have been distracting and that the preschoolers' capacity to respond to verbal incentives may not have been sufficiently developed.

Pollock (1990) compared the effects of two types of examiner comments and three types of modeling experiences on the WISC-R Arithmetic, Picture Completion, Block Design, and Digit Span subtest scores of fourth-grade students. Thirty-nine male and 47 female participants (age range 9 years, 0 months – 10 years, 6 months) were randomly assigned to one of six treatment conditions: (a) modeling videotape with narration and examiner verbal praise; (b) modeling videotape without narration and examiner praise; (c) no modeling and examiner verbal praise; (d) modeling videotape with narration and examiner neutral comments; (e) modeling videotape without narration and examiner neutral

comments; (f) no modeling videotape and examiner neutral comments. The Otis-Lennon School Ability Test, Elementary Level Form R (Otis & Lennon, 1982) was administered to all participants prior to treatment (mean IQ was 110).

The six experimental groups were pre-experimentally matched with respect to sex, age, and intellectual functioning. The examiner verbal praise and neutral comments were issued regardless of the correctness of the participant's response after the first three examinee responses, after every third examinee response following the completion of the first three items, and between the subtests. The examiner verbal praise statements consisted of the following: "That's the way to try," "See how focusing helps you with the task," and "Paying attention has really helped. Keep up the good work on the next part" (p. 68). The examiner neutral comments were: "Now try this," "OK," and "Let's try something different" (p. 69).

Pollock failed to observe significant positive effects for verbal praise, modeling, modeling with narration, or interaction between examiner comments and modeling. Pollock suggested that more emphatic verbal praise (e.g., "That was fine," "Very good; keep up the effort" [p. 82]) might have induced a greater degree of effort than her statements. She also argued that her participants might have perceived the verbal praise as a form of negative feedback. The possibility

that Pollock's sample may have been accustomed to performing at an optimal level without the benefit of external incentives was raised.

Studies Showing Positive Effects for Incentive Conditions

Twenty-four studies were identified that presented varying degrees of positive effects regarding incentives and feedback on examinee performance on cognitive tasks. Bornstein (1968) compared the effects of verbal approval, disapproval, and no-feedback incentives on the WISC Performance Scale administered to 90 third, fourth, and fifth-grade students. Three groups of participants, each corresponding to one of three aforementioned incentive conditions, were pre-experimentally equated through a matched randomization procedure. The groups were approximately equal with regard to their gender ratios. The participants' scores on the California Test of Mental Maturity (Sullivan, Clark, & Tiegs, 1957) served as covariates for the WISC scores. Bornstein's approval condition involved the following statements: "Good!", "Fine!", and "That was fine" (p. 22). Disapproval was conveyed by the comments "I thought you could do better than that" or "That wasn't too good" (p. 22). In the neutral condition, no feedback was offered to participants regarding their performance. One male examiner presented the incentive statements on a noncontingent basis after each participant's response to the first subtest item and between the subtest administrations.

Bornstein reported that the mean WISC Performance IQ for participants in the verbal approval group was significantly greater than the mean IQ scores for participants in both the disapproval, neutral, and no- feedback groups. He also observed that mean scores on the Picture Completion, Block Design, and Object Assembly subtests for the verbal approval participants were significantly greater than the scores for participants in the disapproval group and in the neutral group. In addition, the participants in the verbal approval group performed significantly better on the Picture Arrangement subtest than the participants in the disapproval group. Gender differences were also observed. Male participants in the verbal approval group performed significantly better than males in the disapproval group. Female participants who received verbal praise performed significantly better than females in both the disapproval group and the neutral group.

Similarly, Witmer, Bornstein, and Dunham (1971) used the methodology and participant characteristics (48 male and 42 female third- and fourth-grade students) employed by Bornstein (1968). Selected WISC Verbal (Arithmetic and Digit Span) and Performance (Picture Arrangement and Block Design) subtests served as the dependent variables. The participants who received verbal approval attained a significantly higher mean standard score than the participants in the disapproval group. Although the mean standard score for the participants in the

approval group was not significantly greater than the same score for the neutral group, a directional trend in favor of the verbal approval group was reported. The researchers suggested that the difference between the two groups may have been significant if the complete WISC had been administered.

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 participants were initially tested according to the standard directions. On the basis of these scores, the participants were assigned to one of the three groups and were subsequently retested after an interval of approximately three weeks. The praise condition involved the following examiner exclamations which were contingent on correct responses: "Good," "Fine," "Right," "Very good," "Excellent," and "Okay" (p. 874). 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 was defined as the absolute time score. Differences in the scores from pretest to posttest were analyzed. The accuracy difference scores of male participants in the token reinforcement condition were significantly higher than the scores of males in either the social reinforcement or the control groups. Female participants in the social reinforcement condition, however, attained significantly higher scores than females in either the token reinforcement or the control group. Males in the social reinforcement condition took significantly less time to complete the designs upon retesting than males in either the token reinforcement or the control group. A sex by treatment interaction for speed was also noted. In the posttest, males in the social reinforcement condition performed at a significantly faster speed than females in the same treatment.

In a similar vein, Sweet and Ringness (1971) investigated the performance of 156 elementary school male students (age range 6 years - 13 years) on the WISC Verbal Scale under either monetary reinforcement contingent on correct answers, feedback on correct responses, or standardized conditions. Participants had been referred for psychological testing to school psychologists during the previous year. Full Scale WISC IQ scores were available and used to select participants

(range 80 - 100). Participants were described as lower-class whites, lower-class blacks, and middle class whites based on three of the four Indices of Social Characteristics (Warren, Meeker, & Eells, 1960). Information on the participants' father's occupation, house type, and residential area were available, but family's source of income not obtained. Within each socio-economic group, participants were randomly assigned to one of the three treatment groups.

The participants in the monetary reinforcement condition were given the following instructions:

> For each question answered correctly, I am going to give you either a blue or red poker chip. When we are through, you will turn in all the chips you have earned in exchange for money which will be yours to keep ... If your answer is correct, you will receive a blue chip (1 cent); if it is mostly correct, a red chip (1/2 cent) ... it is expected that there will be some questions for which you will be unable to give a correct answer. When this occurs, no chips will be given. (pp. 402-403). The instruction for the participants in the feedback condition were: I am going to tell you whenever you give a correct answer. When your answer is completely correct, I will tell you 'all correct'. When your answer is mostly correct, I will tell you 'mostly correct'. Here is a paper and pencil for you to keep track of your correct and mostly correct answers. (p. 402).

The participants in the standardized condition were instructed to "Try to answer all question to the best of your ability" (p. 402).

The lower-class white participants in the monetary reinforcement group performed significantly better than the lower-class white participants in the feedback and standardized conditions on the WISC Verbal Scale. The participants from this socioeconomic group who received feedback performed significantly better than the participants in the standardized condition. As hypothesized by the investigators, there was no significant difference between treatment conditions for middle-class white participants. In contrast, the hypothesized advantage for lower-class black participants in the monetary reinforcement condition was not demonstrated.

Racial differences and social class differences on test performance under contingent praise, contingent material reward, and standardized administration were studied by Wienges in 1971. Participants were 120 parochial school students in grades 4 through 6 who were divided equally according to sex, race (black and white), and social class (middle and lower). Wienges' criterion for the selection of participants was a score of 90 or higher on the Otis Quick Scoring Mental Ability Test, Forms Alpha and Beta (Otis, 1954). The participants were randomly assigned to one of three treatment conditions and they were

administered an abbreviated form of the WISC. The participants in the verbal praise treatment were commended for correct responses with the statements, "That's good," "You're doing well," or "Right" (p.27). The participants in the material reinforcement condition, on the other hand, received candy, charms, or baseball cards for correct responses. The participants in the 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 participants made significantly greater gains than white middle class participants in the material reward condition. Similarly, white lower class participants demonstrated significantly greater gains in the material reward condition than they did in the verbal praise treatment. Again, black and white lower class participants made significantly greater gains than white middle class participants in the material rewards condition. Weinges concluded that material rewards were more effective than verbal praise and standardized administration.

The effects of reinforcement on the intelligence test scores of children (age range 5 years - 7 years) was studied by Edlund (1972). Eleven pairs of

participants (10 male and 1 female) were identified from a sample of 79 Head Start and Kindergarten students from low-middle class and lower-class homes (method of class determination was not specified). Participants were matched on the basis of Stanford-Binet, Form L test scores, age, sex, a liking of candy, no problem digesting candy, and parent permission to take candy. One participant from each pair was randomly assigned to an experimental reward group and the other was assigned to a control group. All participants were administered the Stanford-Binet, Form M seven weeks after taking Form L by the same school psychologist who had administered the initial test. The participants in the reward group were told: "I am going to give you an M&M candy for each right answer you give to the questions I ask and each thing you do right that I ask you to do" (p. 318). The participants were given the reward immediately after each correct response. The control group was given the test according to the instructions in the manual. All examinations were given just before lunch to ensure a high level of food-deprivation state.

Edlund reported that the reward group demonstrated a significantly greater mean score than the control group on the posttest. Mean scores on the pretest had been identical (82). Furthermore, all but one of the reward participants experienced improved posttest scores while five control participants showed

poorer posttest scores. Edlund identified the importance "that precise reinforcement procedures need to be used in testing procedures if one is to produce an accurate summary of the individual's learning progress or his IQ which may be used as a basis for administrative decisions".

Miller (1974) studied the effects of the type of incentive condition during testing and the participants' level of social deprivation on the PPVT results of 60 institutionalized mentally retarded participants (age range 9 years - 21.5 years). The effects of three types of test incentive conditions (i.e., verbal praise, token reinforcement, and standard procedures) and two levels of social deprivation (i.e., relatively deprived and relatively undeprived) were considered. The dependent variable consisted of raw scores and mental age scores on Forms A and B of the PPVT. Participants were considered to be relatively socially deprived or undeprived based on ranks on the Social Interaction Inventory, a scale specially constructed for this study by the investigator. A high interaction group (relatively undeprived) and a low interaction group (relatively deprived) were identified. The participants were matched for sex, age, IQ, and origin of retardation (organic or familial).

All initial PPVTs were administered according to the test's guidelines. Retesting with an alternate test form for the participants in the verbal praise group

involved examiner praise for correct responses, (e.g., "That's very good," "Hey, another one right," and "Right again; you must do very well in school"). The verbal praise participants also were engaged in conversation for 15 to 30 seconds with the examiner after every third correct response. Participants in the token reinforcement group received a penny for each correct response. The money was subsequently traded for candy. The control participants who received the standard presentation experienced brief, noncontingent approval statements that were not described by the investigator.

Miller reported no main effects on the raw or mental age change scores. An interaction effect revealed that relatively undeprived participants in the verbal praise group attained significantly higher raw scores than the relatively deprived participants. Relatively deprived participants in the token reinforcement group attained significantly higher raw scores than the relatively deprived control participants. Miller suggested that a history of failure in social and academic situations similar to PPVT testing made the test-retest experience aversive for the socially deprived participants.

The effects of different reinforcement contingencies on the cognitive test performance of 44 deaf and hard-of-hearing children (age range 11 years - 18 years) was explored by Smeets and Striefel (1975). The Raven's Progressive

Matrices (Raven, 1960, 1965) was administered to all participants as a pretest measure with reinforcement provided at the end of the test session (a choice of 10 pennies, 10 small candies, or 5 big candies). The participants were divided into four same-size groups that were equivalent with respect to mean ages and pretest scores. Each group was randomly assigned to one of four reinforcement conditions. After a period of 17 days, the participants were administered the Raven's Progressive Matrices by the same examiner who had administered the pretest.

The participants in the end-of-session reinforcement followed the same procedures used during the pretest. The participants in the noncontingent reinforcement group received a checkmark on a special card for every response, correct or incorrect, or at the end of the 20 second response interval if no response was given. In the delayed reinforcement condition, the participants were asked to pause at the end of every sixth trial to receive one checkmark for every correct response made during the preceding six trials. The participants tested in the immediate reinforcement condition received a checkmark immediately after every correct response. At the end of the test sessions, the examiner counted the checkmarks that the participants exchanged for candies or pennies. Data analysis revealed that the mean posttest scores for all four groups exceeded the mean

pretest scores; however, the mean posttest score for the participants in the immediate reinforcement group was significantly greater than the mean posttest scores for the participants in the other reinforcement conditions. Smeets and Striefel concluded that the immediate delivery of checkmarks for correct responses increased test performance, but they could not conclude whether the checkmarks served as reinforcers or merely as feedback.

In a series of three experiments (Saigh, 1981a; Saigh & Payne, 1976, 1979), the effect 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 school-age students of both sexes); instruments (either the full WISC-R or a short form consisting of the Arithmetic, Digit Span, Picture Completion, and Block Design subtests); and use of noncontingent reinforcement. The independent variables were either verbal praise regardless of the correctness of the response, e.g., "That was very good, let's try some more," "Very good," "You're doing well, keep it up," "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 participants. The studies varied in terms of sample size, schedule of reinforcement, and additional treatment variables.

In the earliest study (Saigh & Payne, 1976), four examiners administered a short form of the WISC to 40 educably mentally retarded students (mean IQ = 63.4, age range 7 years - 16 years). Approximately equal numbers of black and white participants of each sex constituted the sample. The participants were randomly assigned to one of two treatment groups. The examiners in the approval condition praised the participant after the first and second items in the WISC Arithmetic, Block Design, Picture Completion, 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 test sessions. They tempered this conclusion with a recommendation for exploring the effects of verbal reinforcement on the Full Scale Score of the WISC.

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 (IQ range 55 - 74, mean age = 11.8 years). In doing so, the participants were randomly assigned to one of six conditions. In the first condition (i.e., fixed ratio-verbal praise), the participants were verbally praised (e.g., "That was very good; keep it up") after the first, second, and third items on each subtest without regard to correctness and between subtests. In the second condition (i.e., continuous verbal praise), the participants 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 (i.e., fixed ratio-token reinforcement), the examiners gave tokens that were exchangeable for chocolate after the first three responses of a subtest and between subtests. In the fourth condition (i.e., continuous token reinforcement), the participants were given exchangeable tokens after every response and between subtests. The fifth condition (i.e., fixed ratio-neutral) involved the use of neutral, non- evaluative examiner comments (e.g., "Let's try this") after the first, second, and third examinee responses 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 Digit Span, Arithmetic, and Picture Completion mean scaled scores of the verbal praise and the token reinforcement groups significantly exceeded the mean scaled scores of the controls. No significant differences were observed as a function of the type of reinforcement that was used. It should be noted that the total amount of reinforcement under both schedules was very close. 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 suggested that the token reinforcement and the verbal praise may have contributed to maintaining the participants' attention and concentration that were called for on the Arithmetic, Digit Span, and Picture Completion subtests. With regard to the Block Design subtest which involved analysis, synthesis, and reproduction of abstract figures, 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 11 subtests) is a critical factor in placement decisions of special education students, Saigh (1981a) investigated the effects of positive reinforcement during the administration of the Full Scale WISC-R. In this investigation, 22 black and 18 white educably mentally retarded students (mean age = 11.5 years, mean IQ = 72, equal numbers of males and females) comprised the sample. The participants 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 examiners 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 participants' Verbal, Performance, and Full Scale IQ scores. He also noted that on five out of 11 WISC-R subtests (Vocabulary, Arithmetic, Picture Completion, Digit Span, and Coding) the mean scaled scores of the verbally praised participants significantly exceeded the scores of the controls. Saigh suggested that the examiners' positive comments may have reduced anxiety, facilitated attention and concentration, or induced effort among the participants. These effects may have interacted with the unique mental-operation demands presented by the various WISC-R subtests to

account for the differential treatment effects observed on the subtests. Analogously, verbal praise may have motivated the participants to become more verbal or to generate self-instructions that matched the examiner praise.

The effects of individualized incentives on norm-referenced IQ test scores were investigated by Breuning and Zella (1978). The participants were 485 high school student from 19 special education classes at seven suburban Chicago high schools. It should be noted that students diagnosed as having reading handicaps were excluded from the study. Equal distributions of males and females and freshmen, sophomores, juniors, and seniors were reported. All participants had been tested (generally in their freshman year) on the Otis-Lennon Mental Ability Test (129 participants), the Lorge-Thorndike Intelligence Test (147 participants), or the WISC or the WISC-R (209 participants) to obtain diagnostic IQ scores. The participants' Incentive Conditions (46 in each class) were randomly assigned to either the control group or the incentive group. Analyses of the Verbal, Performance, and Full Scale IQ scores for each test revealed no significant differences between scores for the two groups. Analogously, no differences between groups were found with regard to sex, class standing, ethnic status, and socioeconomic status (method for determining socioeconomic status was not identified). One hundred and eighty-nine of the control group participants and

184 of the incentive group participants were retested on the same IQ test by the same trained school psychologist who had administered the initial diagnostic test. The mean pretest-posttest interval was 21.6 months (range 4.3 months - 34.6 months). The participants in the control group were retested according to instructions in the test manual. In addition to the standard instructions, the participants in the incentive were told:

I want you to retake this test. As soon as you finish, the test will be scored. If you have done a lot better than you did the first time you took the test [sic] you can have your choice of the following: (student-specific incentives).

For each participant in the incentive group, several potential incentives were selected based on interviews conducted with the participant and the participant's friends, parents, and teachers. Similar interviews were conducted with the control participants and their friends and parents to reduce procedural bias. The most frequently identified maximal incentives included record albums, trips to professional sporting events, aquarium set-ups, and portable radios. The incentives were generally presented in quantities greater than one, and the cost did not exceed $25.00.

Breuning and Zella reported that the mean WISC-R posttest Verbal, Performance, and Full Scale IQ scores on the three instruments for participants in

the incentive group were significantly higher than the scores for the participants in the control group. An increase of approximately 17 points was reported. Extreme scores and subtest analyses were examined and were not found to cause the overall increase for incentive group participants.

In an adjunct study, Breuning and Zella replicated the procedures used in their study involving special education students with two groups of non-special education students. One group of 200 participants had pretest IQ scores between 98 and 120. The other group of 100 participants had IQ scores between 121 and 140. The incentive group participants in the 98 - 120 IQ group showed posttest increases on Verbal, Performance, and Full Scale scores averaging between 3 and 7 points. The control group participants had increases between 1 and 4 points. The incentive group and control group participants in the 121 - 140 IQ group showed posttest increases between 1 and 5 points. In comparison with the results of special education participants, the incentive condition effects were much less pronounced for non-special education participants with pretest IQs between 98 and 120 and did not occur for participants having pretest IQs between 121 and 140. Breuning and Zella concluded that many students perform poorly on norm-referenced tests because they are not motivated to perform at their maximal level.

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 socioeconomic black second grade students (the gender composition and basis for social class determination were not disclosed) were randomly assigned to one of four treatments and were tested with a short form of the WISC-R (an unspecified combination of subtests proposed by Silverstein, 1967). The senior author, a black doctoral level psychologist, was the examiner. The scoring was done by a master's level clinician who was unfamiliar with the participants and the purpose of the study. The participants in the non-reinforcement group were tested according to manual procedures. In the candy reward condition, the participants were reinforced with an M&M after each correct response. The participants in the traditional social reinforcement group were rewarded for their correct responses by remarks such as "Good" and "Fine" (p. 1538). The participants in the culturally relevant social reinforcement group, on the other hand, were congratulated for correct responses with praise such as "Good job, blood," and "Nice job, little brother" (p. 1538). The report did not specify a description of attempts to ensure that the four groups were equivalent.

The results showed no significant differences between the control and the traditional social reward groups, nor between the social reward and the candy reward groups. The mean score of the participants in the candy reward condition, however, significantly exceeded the score of the control group participants. Finally, the participants in the culturally relevant social praise group obtained a mean score that was significantly higher than scores of the participants in the control and traditional reinforcement groups. The researchers concluded that the type of social reinforcer has an important effect on black children's performance on cognitive tests.

In a subsequent study involving 120 black male elementary school students (age range 9 years - 11 years), Terrell, Terrell, and Taylor (1980) contrasted nonreinforcement, tangible reinforcement, traditional social reinforcement, and culturally relevant reinforcement following the same procedures used by Terrell et al. (1978). A second independent variable (the race of the examiner) was added. Again, the investigators did not provide a description of attempts to ensure that the four groups were equivalent. No significant main effect for the race of examiner was detected. Terrell et al., however, noted a main effect for the type of reinforcer: the participants given the tangible rewards (i.e., candy) obtained significantly higher mean IQ scores than the participants given no reinforcement

or the participants given the traditional social reinforcement. Two interaction effects were also reported. The participants in the culturally relevant reward condition with a black examiner obtained significantly higher scores than the participants in the same experimental situation with a white examiner; the participants who were tangibly rewarded by the white examiner performed significantly better than the participants in the other experimental conditions with the white examiner. Terrell et al. concluded that a tangible reward was the most effective reinforcer for white examiners to use in raising the intelligence test scores of black participants, whereas tangible and socially relevant reinforcers were equally efficacious when administered by black examiners to black participants.

The third study by Terrell, Terrell, and Taylor (1981) involved 100 mildly retarded black-male special education students (age range 9 years - 11 years). In this investigation, IQ scores on the WISC-R or Stanford-Binet Form L-M used for special education placement served as the pretest scores. It should be noted that the distribution of the Binet scores in the experimental conditions was determined, and the range was not considered to be great enough to have affected the results significantly. The participants were randomly assigned to one of four experimental conditions in which they subsequently received a different

administration of the complete WISC-R. The participants 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," "Nice job, little brother," and "Good work, young blood"). Data analysis revealed significant differences between groups. Culturally relevant social reinforcement and the tangible reward treatments were found to be equally effective in terms of raising the participants' test performances 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 students labeled as mentally retarded. Moreover, the authors observed that the gain scores of the participants 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.

Willis and Shibata (1978) compared the effects of tangible reinforcement and feedback on the WPPSI IQ scores of 30 rural Canadian preschool children (20 males and 10 females, age range 3 years, 3 months – 6 years, 6 months). The participants attended a publicly supported day nursery which served lower

socioeconomic families who were referred to the program by private and public social service agencies. All participants were pretested on the WPPSI under the standard procedure outlined in the manual that included several pre-experimental play contacts to establish rapport. Verbal praise (e.g., "You did very well") for the participant's effort, regardless of accuracy, was provided at the completion of each subtest and at points designated on a specially prepared record form during testing. After the pretest, the participants were divided into three groups balanced for sex, age, and pretest IQ score. The participants in the control group were retested on the WPPSI under the standard procedure which was identical to that used for the pretest. The participants in the feedback condition were retested in the same manner with the addition of a means of informing the participant when items were answered correctly. The participants were told:

We are going to play the same game we did once before. One thing is different this time. Whenever you give me a correct answer, you get one of these little pegs which I will stick into there (a wooden board with holes placed in front of the child). The more correct answers you give me (or good work), the more pegs you get to stick into here. Now, let's try some practice questions to see what we are going to do (p. 35).

If a participant missed the examiner praise because of an error on the designated item, praise was given on the next correct item to ensure that all treatment groups received an equal amount of praise. The accumulated pegs were collected and counted four times during the test session. The participant was allowed to walk about the room to approximate the time spent for the reinforcement procedure in that condition. The participants in the reinforcement condition were retested in the an identical manner to participants in the feedback group; however, the participants' accumulated pegs were exchanged for back-up reinforcers (typically worth one dollar) which included candies, small toys, marbles, balloons, and pennies. The average pretest to retest interval was 16.3 days.

Willis and Shibata's results showed that the mean WISC Verbal, Performance, and Full Scale IQ scores for participants in the reinforcement group were significantly greater than the scores for the participants in the control and feedback groups which were not found to differ significantly from each other. The reinforcement group consistently performed better than the other two groups on all subtests. Significant differences between the reinforcement group and the control or the feedback groups were shown on the Vocabulary, Arithmetic, Picture Completion, and Geometric Design subtests. The results of this study

helped to clarify the problem created in studies in which the confounding of feedback and tangible reinforcement occurred.

The influence of verbal praise on the WISC-R performance of intellectually superior elementary and secondary students was investigated by Saigh (1981b). Twenty-four white students (13 males and 11 females) enrolled in a private American overseas school were selected randomly from a pool of 40 students (average Lorge-Thorndike IQ = 124, average age 13.5 years). The participants were randomly assigned to one of two treatment groups and were administered the complete WISC-R. The participants in the examiner praise group were told "Good, keep up the fine effort," "I like the way you're working," and "That's the way!" (p. 647) after the first, second, and third items on each subtest regardless of the accuracy of the answers. The participants were similarly praised after every other response beyond the third item on each subtest and between subtests. In the same manner, the participants in the neutral feedback group were told "Now try this one...," "OK, answer this...," and "Give this one a try..." (p. 647). The mean Verbal, Performance, and Full Scale IQ scores and the mean scores on 5 of the 11 subtests for the examiner praise group were significantly greater than the scores for the neutral feedback group. Saigh (1981b) concluded that the

examiner's verbal praise may have facilitated a substantial level of task-related effort.

Kieffer and Goh (1981) evaluated the effect of individually contracted incentives on intelligence test performance of 96 middle-and low-socioeconomic status (SES) third-and fourth-grade students living in rural communities. One half of the participants were from low income homes as indicated by their eligibility for a federally funded free hot lunch program which required that the family income of the student be less than the State designated poverty level. The family incomes of the middle- socioeconomic group were considerably higher than those in low income homes. Potential individually contracted rewards, grouped into two classes of incentives (tangible and social), were determined on a revised Mediator-Reinforcement Blank (Tharp & Wetzel, 1971). Approximately equal proportions of participants within each of the middle- and low socioeconomic groups indicated a preference for similar types of either tangible or social incentives. Typical social incentives included playing a game with a parent at home, riding a horse or bicycle, and staying all night at a friend's house. Typical tangible incentives included money (usually 25 cents), candy, ice creme, and gum. The Quick Test (Ammons & Ammons, 1962) was used as a screening

instrument to insure that all participants were in the normal range of intellectual functioning (IQ = 90 - 110).

The participants were randomly selected and assigned to six conditions: middle SES - tangible incentive, middle SES-social incentive, middle SES - control, low SES -tangible incentive, low SES - social incentive, and low SES - control. All participants were administered a short form of the WISC-R comprised of two verbal subtests (Information and Similarities) and two performance subtests (Picture Completion and Block Design). The participants in the four SES X reward treatment conditions received the following instructions:

> I am going to ask you a number of questions and to put together some puzzles, and I want to see how good you can do them. You said that you would be willing to work for (child's preferred reward) and if you try to do your best for the next ten minutes on this game, I can see that you receive the reward you want the most after school from your parents. (At this time the examiner and child established the reward the child will be working toward.) You have to try to answer all the questions to the best of your ability. Is it a deal? (Five-second pause). Good, let's begin (pp. 177 -178).

The participants in the two control groups received the standard administration and instructions specified in the WISC-R manual. The obtained results on the

WISC-R short form were transformed into Full Scale Deviation IQ scores using procedures suggested by Tellegen and Briggs (1967). Data analysis revealed a significant main effect between SES mean scores in favor of the middle- class participants. No significant main effect due to reward treatment condition was found. A significant interaction effect between SES and reward conditions was reported. The differences between middle- and low- SES mean IQ scores in both the tangible and social reward conditions (6.5 and 3.9 respectively) were significantly smaller than the difference between mean IQ scores in the control condition (17.2). Individually contracted incentives based on careful assessment were shown to be effective in increasing the IQ scores of low- SES children.

The results indicated that a significant portion of the observed IQ difference between middle- and low- SES children may be accounted for by the motivational deficit usually experienced by low- SES participants during standardized testing conditions. Standard testing procedures appeared to be sufficient in motivating the middle- SES participants as much as tangible or social incentives did.

A series of six studies investigating the effects of token reinforcement on the WISC-R performance of groups with varying characteristics shared the same investigators and similar methodology (Bradley-Johnson, Johnson, Shanahan, Rickert, & Tardona, 1984; Bradley- Johnson, Graham, & Johnson, 1986; Devers,

Bradley- Johnson, & Johnson, 1994; Galbraith, Ott, & Johnson, 1986; Johnson, Bradley-Johnson, McCarthy, & Jamie, 1984; Young, Bradley-Johnson, & Johnson, 1981).

In the first study, Young et al. (1982) compared the effect of immediate reinforcement against delayed reinforcement on the WISC-R performance of 30 white mildly mentally retarded students (IQ range 50 - 70). The participants included 19 males and 11 females (age range 7 years - 12 years) who were from low to middle social class families based on Hollingshead's Two Factor Index of Social Position. All participants had been administered the WISC-R on routine evaluations at least 6 months, but no more than 3 years, prior to the date of this study. Reinforcement procedures had not been used with the participants during testing, nor were they on token economies in their classrooms.

The participants, who were randomly selected from a pool of 34 participants, were randomly assigned to one of three treatment conditions: control, delayed reinforcement, and immediate reinforcement. The three treatment groups were found to be equivalent with regard to mean age and mean IQ scores on the Slosson Intelligence Test (Slosson, 1963) which was initially administered to all participants. Subsequently, the 10 regular subtests of the WISC-R were administered by an experienced school psychologist whose performance was

participated to two reliability checks for each group. The control group was tested under the standard conditions identified in the WISC-R manual which included social reinforcement for effort (e.g. "You are really working hard today"). This group was tested first so that their performance was not affected by the knowledge that the participants in the other groups received tangible reinforcers. The participants in the two experimental groups were tested in random order in the same manner as the control participants.

Prior to WISC-R testing, the participants in the immediate reinforcement group were told that they could exchange earned tokens for items halfway through the exam and at the end of the test. The participants in the delayed reinforcement group were given the same instructions except they were told that they would receive the tokens halfway through the assessment and at the end of the test. Each response was delegated the same number of chips as the number of points the participant earned for each item. The participants earned between 50 and 120 tokens which were exchanged for a variety of back-up reinforcers which were chosen from a list generated when the examiner asked the participants what kinds of things they liked that cost less than $2.00. The items included: raisins and pens (10 tokens), crayons, hair ribbons, and trucks (30 tokens), coloring books and rings (50 tokens), storybooks (70 tokens), and games (100 tokens).

The mean WISC-R Full Scale IQ scores for the participants in the immediate reinforcement and the delayed reinforcement groups were significantly higher than the score for the participants in the control group. No differences were found between the two reinforcement groups. On the Verbal Scale, the mean score for the participants in the immediate reinforcement group was significantly higher than the score for the control group participants. No differences were observed between the delayed reinforcement and control groups nor between the two reinforcement groups. No significant difference was found between the groups on the Performance Scale scores. Delayed reinforcement seemed to be as effective as tangible reinforcement in enhancing the overall WISC-R performance of mentally retarded children. Young et al. concluded that many mentally retarded children are not maximally motivated to perform under standardized testing conditions. To obtain an accurate assessment of a child's capabilities, both acquired skills and the child's motivation to perform the tasks must be considered.

In a similar vein, Bradley-Johnson, Johnson, Shanahan, Rickert, and Tardona (1984) replicated Young et al.'s (1982) study by conducting two experiments which involved urban, black and white, low SES second- grade students (sex distribution was not reported). All participants came from low SES families based on Hollingshead's Two Factor Index of Social Position. In the first study,

33 black participants were randomly selected from 35 children in a second-grade class and assigned to one of three treatment conditions: control, delayed reinforcement, and immediate reinforcement. The same procedures used in the Young et al. study were followed with several exceptions noted. The participants in the reward conditions were shown the reinforcers prior to their being tested and exchanged their tokens for rewards at the end of the test administration. The participants in the delayed reinforcement group were given their tokens at the end of each subtest. Three white examiners (two males and one female) administered the entire WISC-R except the Maze subtest. The immediate reinforcement group scored significantly higher than delayed reinforcement and standard administration groups on the mean WISC-R Verbal, Performance, and Full Scale IQ scores. The performance of the delayed reinforcement group did not differ significantly from the standard administration group.

In the second study, 33 white participants were serially selected from 52 second-grade children as permission slips were returned by the participants. The procedures were exactly the same as those used in the first experiment. Two white-male examiners administered the WISC-R. No significant differences were shown between conditions on the mean Verbal, Performance, and Full Scale scores. Based on their results, Bradley-Johnson et al. suggested that the effect of

contingent token reinforcement for correct responses on children's IQ scores would differentially affect different subgroups.

Johnson, Bradley-Johnson, McCarthy, and Jamie (1984) continued to study the effect of token reinforcement WISC-R performance in two experiments that involved black, educable mentally impaired students. In the first experiment, 20 elementary-age (age range 6 years, 6 months - 12 years, 8 months) students from three rural schools were selected serially as their parent permission slips for participation in the study were returned. All of the participants' families were assessed to be of low socioeconomic status on Hollingshead's Two Factor Index of Social Position. The participants were randomly assigned to one of two groups, the standardized administration group or the standardized administration plus token reinforcement group. The procedures used in this study were similar to the procedures used by Young et al. (1982) and Bradley-Johnson et al. (1984) for their control and immediate reinforcement groups. One except was noted. The Slosson Intelligence Test (Slosson, 1963) was not administered to ensure that participants from both groups were equivalent on intelligence test scores; however, all participants were identified as classified educable mentally impaired students. The WISC-R was administered by an experienced white male school psychologist. The results revealed significant differences on the mean Verbal and

Full Scale IQ scores (13.3 and 8.9 points respectively) for the participants in token reinforcement group. No significant difference between groups was observed for the Performance Scale.

In the second experiment, 22 junior high school age (age range 12 years, 7 months - 14 years, 11 months) students from two schools were selected in the same manner reported in the first study. All of the participants' families were assessed to be of low socioeconomic status on Hollingshead's Two Factor Index of Social Position. The participants were randomly assigned to either the standardized administration group or the standardized administration plus token reinforcement group. The procedures used in this experiment were identical to the procedures used in the first study. It should be noted that the back-up rewards chosen by this older group of participants included hair picks, restaurant coupons, records, and candy (all costing less than $3.00). In addition, the examiner was a white female school psychologist. No group differences were found on mean Verbal, Performance, or Full Scale IQ scores. The use of token reinforcement did not influence the WISC-R test performance for the older black students classified as educable mentally retarded to the same degree that the younger group was influenced. The authors concluded that some groups of children need token

reinforcement more than others to motivate them to perform optimally on standardized tests of intellectual ability.

Bradley-Johnson, Graham, and Johnson (1986) expanded their study of the effects of token rewards on the performance of the WISC-R which involved low-socioeconomic status, upper and lower elementary school-age students. Their participants were 40 white, rural students (19 males and 21 females). Half of the participants were in the first and second grades, and half were in the fourth and fifth grades. The families of all participants were determined to be in the low-socioeconomic status group on Hollingshead's Index. The participants in each age group were randomly assigned to one of two experimental conditions, a token reinforcement group or a standard administration group.

A well qualified, white female examiner administered the WISC-R to the participants. The procedures were identical to the procedures used by Bradley-Johnson et al. (1984) with their control and immediate reinforcement participants. It should be noted that the four groups were equivalent on intellectual functioning that was assessed prior to the treatments with the Slosson Intelligence Test for Children (Slosson, 1975). Data analysis showed that the mean Verbal, Performance, and Full Scale IQ scores for participants in the token reinforcement

group were significantly higher than the scores for control group participants. No significant interaction effects with age were obtained.

The methodology used by Bradley-Johnson et al. (1984) and Bradley-Johnson et al. (1986) was employed by Galbraith, Ott, and Johnson (1986) to study the effects of token reinforcement on the WISC-R performance of Hispanic second-grade students. Thirty bilingual children with Hispanic surnames whose primary language was English were selected from English speaking classrooms in one of two elementary schools in an urban, low-income area. All participants were performing satisfactorily in regular education classrooms. The participants were identified as low- socioeconomic status based on the report that they were from low-income families. The participants had not been previously tested with token reinforcement procedures and had not been on token economies in school.

The participants were randomly assigned to two groups each composed of eight males and seven females (respective age ranges 7.6 years - 9.5 years and 7.4 years - 9.4 years, respective mean ages 8.5 years and 8.3 years). One group was administered the WISC-R under standard conditions, and the other group received token reinforcement for correct responding. The WISC-R was administered by two well trained and experienced Caucasian male examiners. In addition to implementing Bradley Johnson et al.'s procedures, Galbraith et al. assessed the

effectiveness of the tokens used with the reinforcement group during three (ABA) 30 second trials of a bead stringing task. During the B (reinforced) trial, all participants exhibited higher rates of work as they earned a token for each bead strung. Tokens were exchanged for participant-selected back-up reinforcers.

Six reliability checks revealed high levels of treatment integrity. Both groups performed equally on the pre-treatment Slosson Intelligence Test (Slosson, 1975). The token reinforcement group scored significantly higher on both the Full Scale IQ (11 points) and the Performance Scale (13 points) and marginally higher on the Verbal Scale (8 points) than the standard administration group. The investigators concluded that token reinforcement can increase the IQ scores of low-income, second-grade Hispanic children on the WISC-R. This increase was attributed, in part, to increased motivation on the part of the reinforced participants.

The final study in a series of similar investigations of the effects of token reinforcement on WISC-R performance involved 25 Native American students from the Chippewa tribe (Devers, Bradley-Johnson, & Johnson, 1994). The participants were enrolled in fifth through ninth grade regular education classrooms. All but one participant attended a school system serving a large number of Native American students. One participant attended a Catholic school.

Initially, 52 students were named by the Tribal Education Office for possible inclusion in the study. The investigators were able to obtain permission for 31 participants (25 through visits to homes and tribal offices after a first mailing produced only six permission slips).

The 31 participants who were randomly assigned to either the control (standardized administration) or the experimental group (standardized administration, plus token reinforcement for correct responses) were reduced to 25 participant through attrition. The final experimental group was composed of 6 females and 6 males, and the control group consisted of 10 females and 3 males. The WISC-R was administered to all participants by an experienced Caucasian male examiner. Devers et al. implemented the same procedures used by Young et al. with their control and immediate reward groups (participants were able to purchase back-up rewards at the midpoint and end of testing). In addition, the participants in the experimental group were administered two subtests (Word Opposites and Symbolic Relations) from the Detroit Test of Learning Aptitude-2 (Hammill, 1985; DTLA-2) to determine the effectiveness of the tokens and back-up rewards and to expose the participants to reinforcement contingencies.

Higher mean standard scores on the DTLA-2 subtests were observed for participants who received tokens and back-up rewards for correct responses.

Three reliability checks revealed high levels of treatment integrity. Both groups performed equally on the pre-treatment Slosson Intelligence Test (Slosson, 1975). The token reinforcement group scored significantly higher on the Full Scale IQ (12 points) than the standard administration group. Devers et al. concluded that token reinforcement can enhance the IQ scores of Native American students on the WISC-R. In addition, the higher mean Verbal and Performance Scores (13.5 points and 8.7 points respectively) of the experimental group compared to the control group indicated that Native American students are capable of providing more information than they typically provide under standard testing conditions.

Incentive Studies Involving Emotionally Disturbed Behavior Disordered Youth

Three studies that examined the effects of alternative test administration conditions on the WISC-R test performance of emotionally disturbed, behavior disordered youth are particularly relevant to this present study.

Mayfield (1977) paired nonverbal behaviors (e.g., smiles and nods) with social reinforcement (non-contingent positive verbalizations). She compared this combination with the presence or absence of a prior relationship (defined as a ten minute "warm-up" before test administration that included informal play and conversation) between the examiner and the participant. The participants were 24 male (15 black and 9 white) and 24 female (11 black and 13 white) clients at a

psychoeducational center (age range 7 years - 11 years). The participants were selected on the basis of average performance on the Slosson Intelligence Test (Slosson, 1963) and extreme scores on the Quay-Peterson Behavior Problem Checklist (Quay, 1967).

All participants were randomly assigned to one of four experimental conditions: social reinforcement and prior relationship with the examiner, social reinforcement and no prior relationship with the examiner, no reinforcement and prior relationship with the examiner, and no reinforcement and no prior relationship with the examiner. The gain scores on the WISC-R Performance subtests comprised the dependent variables. 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 the test administration.

Mayfield reported that the mean scores of participants who were reinforced significantly exceeded the scores of those who were not reinforced. In addition, the participants in the prior relationship-reinforcement group and the participants in the prior relationship-no reinforcement group achieved significantly higher

scores than those in the no 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.

Jackson, Farley, Zimet, and Gottman (1979) studied the effects of alternative test administration conditions on the WISC-R test performance of emotionally disturbed students whose diagnoses ranged from moderate reactive conditions to severe neurotic and psychotic conditions. Seventy-five males and 26 females (age range 6 years-16 years) identified as having behavioral and emotional difficulties based on psychiatric interviews and psychological testing completed the Porteus Maze Test (Porteus, 1942) to obtain a rating of impulsivity. Each participant was randomly assigned to one of the five WISC-R administrative conditions that were explained to the participants in their respective groups before beginning the test. In the reinforcing attention condition, the participant was rewarded after each item with his or her choice of candy (offered at the end of testing) or a penny when the participant responded attentively by answering positively to one of the following questions: "Are you ready?", "Are you paying attention?", or "Are you with me?" (p. 58). In the reward for success condition, the participant was rewarded after each item with candy or a penny for correct answers. In the self vocalization

condition, the participant was requested to read aloud before each item the following statement written on a visual reminder card which was placed on the desk on front of the child: "I will stop, listen, look, and think before I answer" (p. 58). In the feedback on success and failure condition, the examiner indicated if the answer given for each item was correct or incorrect. In the standard condition the participant was treated according to the procedures outlined on the WISC-R manual.

Jackson et al. found no significant difference among the five administrative conditions in an overall analysis that combined all participants and did not consider gender differences or different levels of impulsivity.

In a second analysis of the data, Jackson et al. found significant main effects for administrative conditions, sex, and impulsivity. The low-impulsive participants performed significantly better than the high-impulsive participants on WISC-R Verbal, Performance, and Full Scale scores. The female participants achieved significantly higher Full Scale scores than male participants, and male participants achieved significantly higher Performance scores than female participants. Pair-wise comparisons of means revealed significant differences between administrative conditions and established ranking of administrative conditions that led the investigators to conclude that there is a strong relationship

between WISC-R test scores and administrative procedures that may be unique to children with specific characteristics. Emotionally disturbed boys whose diagnoses ranged from moderate reactive conditions to severe neurotic and psychotic conditions and high-impulsive children performed best under conditions that provided knowledge of success and material rewards or payment for desired behaviors. Emotionally disturbed girls whose diagnoses ranged from moderate reactive conditions to severe neurotic and psychotic conditions and low-impulsive children performed best when given feedback on success or failure on their performances. Jackson et al. suggested that further research should be conducted involving participants who are more homogeneous with regard to age and psychiatric classifications.

Saigh (1981a) identified the need to conduct research on the effects of verbal mediators on the test performance of intellectually divergent groups including emotionally disturbed individuals. He also recommended that studies involving the validation of examiner verbal praise procedures during testing occur. In this vein, Saigh (1982) randomly assigned 51 institutionalized, adolescent females with conduct disorder to one of three treatment groups. The participants in the examiner praise group were told "That was good" (p. 771) contingent on the number of items that were attempted. The participants in the token reinforcement

group were placed on the same contingency. The participants in the neutral feedback group were told "Now try this" (p. 771). A short form of the WISC-R which included the Information, Arithmetic, Digit Span, Picture Completion, and Block Design subtests was administered to the participants. The mean scaled scores for the token reinforcement group on the Digit Span, Picture Completion, and Block Design subtests were significantly greater than the mean scores for the neutral feedback group. In comparisons between the examiner praise and the neutral feedback groups and the token reinforcement and the examiner praise groups, nonsignificant differences between the mean scaled scores were noted across measures. Saigh suggested that material incentives may be a viable means of facilitating the WISC-R performance of adolescents with conduct disorder; however, no significant correlations were observed between the mean scaled scores on the WISC-R subtests and the mean semester grade point averages for all treatment groups.

Summary of Effects of Incentive in Testing

Examined in toto 44 studies were reviewed. Of these, four reports (Benton, 1936; Hurlock, 1924, 1925; Klugman, 1944) were included because of their historical importance in introducing incentive conditions as an independent variable in cognitive assessment. Of the remaining 40 studies, 27 provided

varying degrees of support for the proposition that incentives facilitated performance. Thirteen investigations failed to show a significant effect. Many studies had one or more methodological weakness. The absence of a control group was noted in the following studies: Klugman, 1944; Cohen, 1970; Quay, 1971, 1975; Smeets and Striefel, 1975; Tufano, 1976. Eight studies (Tiber & Kennedy, 1964; Cohen, 1970; Quay, 1971, 1975; Wienges, 1971; Galdieri et al., 1972; Terrell et al., 1978; Terrell et al., 1980) did not provide sufficient information regarding the initial equivalence of the groups.

Many studies showed either no effects or contradictory results with regard to incentive conditions during cognitive testing. Graham (1971) suggested that when participants' level of motivation is sufficient, examiner incentives above and beyond standard procedures may have little effect on test results. She indicated that the individual attention that participants receive may establish a sufficient level of motivation for some individuals. Quay (1971) hypothesized that participants performing close to their intellectual limit may not be as responsive to incentives during testing as participants who perform below their optimal performance level. Investigators who showed that incentives did not improve scores for white middle class participants (Sweet & Ringness, 1971; Weinges, 1971) and participants who performed in the above average and superior range of

intellectual functioning (Pollock, 1989 and Breuning & Zella, 1978) supported Quay's hypothesis. Pollock (1989) speculated that verbal praise may need to be emphatic to show the positive effects of that type of test performance incentive. She also pointed out that positive attention and praise are not always perceived as desirable by participants. In a similar vein, Goh and Lund (1977) suggested that excessive verbal feedback may have been distracting for young participants.

Studies deemed adequate included: Bradley-Johnson et al. (1984), Bradley-Johnson et al. (1986), Bornstein (1968), Edlund (1972), Galbraith et al. (1986), Goh and Lund (1977), Keiffer and Goh (1981), Mayfield (1976), Pollock (1989), Saigh (1980,1981a), Saigh and Payne (1976,1979), and Witmer et al. (1971). Elements which contributed to the worthiness of these studies were: treatment groups equated for age, race, sex, socioeconomic status, and intelligence; examiner reliability checks; precision in identifying behavior to be rewarded; and selection of participant-relevant reinforcers.

Verbal praise by the examiner and tangible rewards had a positive effect upon the cognitive performance of diverse groups: elementary school children, junior high school and high school students, preschoolers, special education students, regular education students, mentally retarded children, hearing handicapped students, emotionally disturbed and behavior disordered youths, and gifted

students. Four racial or ethnic groups were represented in the studies (blacks, whites, Hispanics, and Native Americans). Many studies identified participants by social class (lower and middle), and in many cases, investigators compared the performance of participants from the two classes under different incentive conditions.

Most of the studies examined the effects of incentives on participants' performance on one of the Wechsler Scales (WISC, WISC-R, or WPPSI), or selected groups of Wechsler subtests. Other instruments used by investigators were the Stanford-Binet (Form L, Form M, and Form L-M), the Lorge-Thorndike Intelligence Test (or it's Vocabulary subtest), the Peabody Picture Vocabulary Test, Raven's Progressive Marices, and the Otis-Lennon Mental Ability Test. In two studies (Saigh & Payne, 1976 and Witmer et al., 1971) the researchers, who had used selected Wechsler subtests, urged that further studies be completed using the full Wechsler Scale.

A host of studies have shown that incentive conditions involving examiner verbal praise and tangible rewards can significantly improve the performance of participants on individually administered intelligence tests. Improvement was shown to occur when incentives were contingent and non-contingent on success and when rewards and praise were administered continuously and intermittently.

Studies that examined the effect of verbal reproof on test performance (Bornstein, 1968; Tiber & Kennedy, 1964; and Witmer et al., 1971) did not find that condition to improve scores. Feedback by examiners was found to have no significant effects (Cook, 1973; Goh & Lund, 1977; and Willis & Shabata, 1978) and mixed effects (Jackson et al., 1979; Sweet & Ringness, 1971) on test performances. The confounding effect of feedback on continuous contingent reward for correct responses was recognized by Smeets and Striefel (1975) and Willis & Shabata (1978).

Many investigators who examined the effects of incentive conditions on low socioeconomic-status and mentally retarded minority students and found the scores of these participants to improve significantly, questioned the accuracy of the initial test scores and the validity of conclusions based on these scores. The problem of comparing the performance of the participants who receive incentives during testing with the normative sample was recognized by some investigators. Similarly, the predictive validity of test results obtained by procedures that include incentives was identified as a problem to be studied.

The application of incentives during cognitive testing was recognized as a procedure that reduced the error variance in test scores associated with a level of motivation that was not optimal. In addition to enhancing motivation and effort

during testing, investigators suggested that the incentives may have improved test performance by facilitating attention and concentration (Saigh, 1981a; Saigh & Payne, 1979), reducing test anxiety (Saigh, 1981a), and enhancing verbal behavior (Saigh, 1981a).

Conduct Disorder Youth

The term conduct disorder refers to a pattern of antisocial behaviors exhibited by children and adolescents which occurs chronically and which impairs the functioning of the individual (Kazdin, 1996). Specifically, the antisocial behaviors include aggressive acts, theft, vandalism, setting fires, lying, truancy, and running away. Children with conduct disorder have long been the focus of psychiatric attention as well as of psychologists, other mental health workers, the police, and the legal system (Cantwell, 1989). Cantwell reported that the earliest child guidance clinics founded in the United States were established to treat children with conduct disorder, and many were attached to local juvenile courts. Within the educational system, special services, teachers, and classes are often provided to manage children with conduct disorder (Kazdin, 1996). Aggressiveness, conduct problems, and antisocial behaviors comprise from one third to one half of all child and adolescent clinic referrals (Kazdin, Siegel & Bass, 1990; Robins, 1981). Conduct Disorder children occur with great frequency among populations who are defined by school psychologists as aggressive, by educators and pediatricians as being learning disabled and hyperactive, and by criminologists as delinquent (Robins, 1991).

Kazdin (1996) identified other factors in addition to the frequency of occurrence that contribute to the magnitude of the social problem that conduct disorder represents. Conduct disorder tends to be relatively stable over time (Robins, 1978), and the condition is associated with a poor prognosis (Wolfgang, Figlio, & Sellin, 1972). According to Huesmann, Eron, Lefkowitz, and Walder (1984), conduct disordered behaviors tend to continue across multiple generations. At present, no clearly effective interventions exist for conduct disorder; although diverse forms of individual and group therapy, behavior therapy, residential treatment, psychopharmacology, psychosurgery, family-focused treatments, and community-based treatments have been implemented (Kazdin, 1996).

Finally, conduct disorder is one of the most costly mental disorders to society in terms of money for services (psychiatric and psychological treatment, family social work, juvenile adjudication and incarceration, special education programs, and social service contacts), financial costs to victims, physical and emotion distress of victims, and the chronic maladjustment and unhappiness of the individual with conduct disorder (Robins, 1981).

Current Nosology

Behavior disorders in childhood and adolescence were first addressed in the context of American psychiatric nosology in the Diagnostic and Statistical Manual of Mental Disorders, Second Edition (DSM-II, American Psychiatric Association, 1968). A set of diagnostic conditions was defined in response to the recognition of disorders of intermediate stability in that age group and included: Hyperkinetic Reaction, Withdrawing Reaction, Overanxious Reaction, Runaway Reaction, Unsocialized Aggressive Reaction, Group Delinquent Reaction, and Other Reaction of Childhood (or Adolescence). The Diagnostic and Statistical Manual of Mental Disorders, First Edition (DSM-I, American Psychiatric Association, 1952) recognized two conditions that were not as specific as DSM-II categories: Sociopathic Personality Disturbance, Dyssocial Reaction and Adjustment Reaction of Childhood, Conduct Disturbances.

In response to DSM-II's weak operational criteria, limited reliability, and modest coverage (Morey, Skinner, & Blashfield, 1986), the American Psychiatric Association (1980) developed the Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III, American Psychiatric Association, 1968) that included not only diagnostic criteria, but also information as to age of onset, associated features, course, predisposing factors, prevalence, sex ratio, and

differential diagnosis. DSM-III identified identified five types of Conduct Disorder: Conduct Disorder, Undersocialized, Aggressive; Conduct Disorder, Undersocialized, Nonaggressive; Conduct Disorder, Socialized, Aggressive; Conduct Disorder, Socialized, Nonaggressive; and Atypical Conduct Disorder. Because DSM-III subtyping for Conduct Disorder lacked clinical utility and consistency with empirical findings (Robins, 1991), the category was revised and replaced in Diagnostic and Statistical Manual of Mental Disorders, Third Edition - Revised, (DSM-III-R, American Psychiatric Association, 1987). Three subtypes of Conduct Disorder were recognized: Group Type, Solitary Aggressive Type, and Undifferentiated Type.

The American Psychiatric Association's (1994) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) presently continues to categorize Conduct Disorder as a disruptive behavior disorder among the conditions usually first diagnosed prior to adulthood. The DSM-IV diagnostic criteria for Conduct Disorder is as follows:

A. A repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate societal norms or rules are violated, as manifested by the presence of three (or more) of the following criteria in

the past 12 months, with at least one criterion present in the past 6 months:

Aggression to people and animals

(1) often bullies, threatens, or intimidates others

(2) often initiates physical fights

(3) has used a weapon that can cause serious physical harm to others (e.g., a bat, brick, broken bottle, knife, gun)

(4) has been physically cruel to people

(5) has been physically cruel to animals

(6) has stolen while confronting a victim (e.g., mugging, purse snatching, extortion, armed robbery)

(7) has forced someone into sexual activity

Destruction of property

(8) has deliberately engaged in fire setting with the intention of causing serious damage

(9) has deliberately destroyed other's property (other than by first setting)

Deceitfulness or theft

(10) has broken into someone else's house, building, or car

(11) often lies to obtain goods or favors or to avoid obligations (i.e.

"cons" others)

(12) has stolen items of nontrivial value without confronting a victim (e.g., shoplifting, but without breaking and entering; forgery)

Serious violations of rules

(13) often stays out at night despite parental prohibitions, beginning before age 13 years

(14) has run away from home overnight at least twice while living in parental or parental surrogate home (or once without returning for a lengthy period)

(15) is often truant from school, beginning before age 13 years

B. The disturbance in behavior causes clinically significant impairments in social, academic, or occupational functioning

C. If the individual is age 18 years or older, criteria are not met for Antisocial Personality Disorder

Specify type based on age at onset

Childhood-Onset Type: onset of at least one criterion characteristic of Conduct Disorder prior to age 10 years

Adolescent-Onset Type: absence of any criteria characteristic of Conduct

Disorder prior to age 10 years

Specify severity:

Mild: few if any conduct problems in excess of those required to make the diagnosis **and** conduct problems cause only minor harm to others

Moderate: number of conduct problems and effect on others intermediate between "mild" and "severe"

Severe: many conduct problems in excess of those required to make the diagnosis **or** conduct problems cause considerable harm to others

(pp. 90 - 91) [1]

The World Health Organization's International Statistical Classification of Diseases and Related Health Problems, Tenth Edition (ICD-10, World Health Organization, 1992), an international nomenclature system for psychiatric conditions, also specifies diagnostic criteria for Conduct Disorder.

Characteristics of conduct disorder may be evident in other diagnostic categories, but critical features differentiate the conditions. Antisocial behaviors are included in criteria for Oppositional Defiant Disorder (major rule violations and serious antisocial acts are absent), Adjustment Disorder with Disturbance of Conduct (Conduct Disorder symptoms occur in response to a recent stressor), Antisocial Personality Disorder (the individual must be at least 18 years old), and

Other Conditions That May Be a Focus of Clinical Attention (reserved for isolated acts in the absence of other diagnostic criteria).

Epidemiology

Epidemiology is the study of the distribution and determinants of health related states and events in populations (Last, 1983). The Institute of Medicine (1989) estimated that the rate of conduct disorder among children from ages 4 to 18 ranged from 2 to 6% (between 1.3 and 3.8 million children in the United States). This rate was similar to Costello's (1989) estimate of 3 to 7% prevalence in the general population. In general, rates of conduct disorder tend to be higher for adolescents (approximately 7% for 12- to 16-year olds) than for children (approximately 4% for 4- to 11-year olds) (Offord, Boyle, & Racine, 1991). Zoccolillo (1993) reported that conduct disorder was 3 to 4 times more common in males than in females. This finding was consistent with a prevalence rate of 6 to 16% for males and 2 to 9% for females that was reported in DSM-IV.

Students with conduct disorder often comprise a majority in classrooms for children classified as behavior disordered and seriously emotionally disturbed. Pullis (1991) reported that between 67 and 76% of the seriously emotionally disturbed students in his study exhibited characteristics of conduct disorder.

Prevalence in regular elementary school classrooms of antisocial and maladjusted behaviors was 4 to 6% (McGhee & Short, 1991). A lower rate among high school students was attributed to the students who exhibited these behaviors leaving or being excluded from school (Cairns, Cairns, & Neckerman, 1989).

Subtypes

Children diagnosed with conduct disorder can vary widely in terms of specific symptoms and associated features of the disorder (Kazdin, 1996). Efforts have been made by different researchers to identify subtypes of youth with conduct disorder whose features, onset, course, response to treatment, and other characteristics vary systematically. One advantage of the investigation of subtypes has been to identify more homogeneous groups that are more likely to respond to specific interventions and procedures than the heterogeneous group of youth with conduct disorders (Kazdin, 1996).

Achenbach (1993) identified aggressive and delinquent patterns of symptoms based on statistical techniques (factor analysis and cluster analysis). Aggressive youths with conduct disorder are likely to engage in fighting, property destruction, and cruelty to other people or animals. This subtype has been shown to be more stable over the course of development. According to Achenbach, delinquent youths are likely to engage in theft, running away, lying, setting fires,

and truancy. These subtypes are not mutually exclusive and overlap with other subtypes.

Patterson's (1982) subtypes, aggressors and stealers, focused on salient symptoms. The focus on salient symptoms was expanded to consider a broader, bipolar dimension of overt (confrontational) and covert (concealed antisocial) behavior (Loeber, Lahey, & Thomas, 1991; Loeber & Schmaling, 1985).

Two types of conduct disorder have been identified (Hinshaw, Lahey, & Hart, 1993). The first type involves childhood onset that can be evident as early as age 5 (DSM-IV) and that may begin with Oppositional Defiant Disorder or Attention-Deficit/ Hyperactivity Disorder. This type is more severe and has a poorer prognosis than adolescent-onset type that is more prevalent. In addition, boys are more likely to exhibit childhood-onset conduct disorder than girls (Kazdin, 1996).

Conduct disorder can coexist with other diagnostic conditions in addition to Attention-Deficit/ Hyperactivity Disorder. Zoccolillo (1992) reported that conduct disorder has been found to be comorid with anxiety and depressive disorders. Conduct disorder may also be associated with Learning Disorders, Communication Disorders, or Substance-Related Disorders (DSM-IV).

The search for specific subtypes of conduct disorder has not been the only approach taken by researchers (Kazdin, 1996). A broader level of evaluating

antisocial behavior and problem behavior theory (Jessor & Jessor, 1977) proposed that problem behaviors cluster together (e.g., delinquent acts, early sexual activity, drug abuse) because they serve similar functions in relation to development.

A progression of type and severity of conduct disorder symptoms over the course of development has been proposed by Loeber and Stouthamer-Loeber (1998) that identified three pathways to boys' problem behavior and delinquency. The overt pathway begins with minor aggression (bullying, annoying others), escalates to physical fighting (individual and gang fighting), and culminates in violence (rape, attack, strong-arm assault). The covert pathway starts with minor covert behavior (shoplifting, frequent lying), is followed by property damage (vandalism, fireseting), and leads to moderate to serious delinquency (fraud, burglary, serious theft). The third pathway, the authority conflict pathway, occurs prior to age 12 and consists of a sequence of stubborn behavior, defiance and disobedience, and authority avoidance (truancy, running away, and staying out late). In the overt pathway, younger children followed the proposed sequence more strictly than adolescents who often appeared to start at the second step. The worst cases engaged in both overt and covert acts. An alternative model involving a single pathway has been proposed by Patterson, Reid, and Dishion

(1992) that began with disobedience, followed by temper tantrums, fighting and stealing.

Risk Factors

A number of risk factors increase the likelihood of the occurrence of the condition (Kazdin, 1996). Major categories of risk factors for conduct disorder include factors related to the child, the parent and family, the school, and peers (Kazdin, 1996; Quay, 1986; Robins, 1991; Short & Shapiro, 1993). Risk factors were reported to frequently co-occur and to interact.

Child-related risk factors for the onset of conduct disorder include difficult temperment based on problematic characteristics such as activity level, emotional responsiveness, quality of moods, and social adaptability (Reitsma-Street, Offord, & Finch, 1985). Neurological deficits and difficulties (in the areas of language, memory, motor coordination, integration of visual and auditory cues, and executive functions of the brain) predict subsequent conduct disorder (Moffit, 1993a, 1993b). Several studies have found that subclinical levels of conduct disorder predict later conduct disorder (Farrington, 1991; Loeber, 1990).

Academic difficulties and lower levels of intellectual functioning have been associated with conduct disorder and often predicted the disorder (Farrington, 1991; Moffit, 1993a); however, Ledingham and Schwartzman (1984) reported

that the opposite relationship also existed (conduct disorder predicted subsequent difficulties in school). Cognitive factors (e.g., negative cognitive response bias, deficient problem solving skills) play an important and well-documented role in antisocial behavior and conduct disorder (Dodge, Price, Bachorowski, & Newman, 1990; Spivack, Pratt, & Shure, 1976).

Parent and family factors that influence the development of conduct disorder identified by Kazdin (1996) included: genetic loading, psychopathology in the family, parent-child interactions, parental separation, divorce and marital discord, birth order (middle children), large family size, and socioeconomic disadvantage.

A review of studies involving monozygotic and dizygotic twins and adoption studies suggests that both genetic and environmental factors combined to place children at-risk for developing conduct disorder. Analogously criminal behavior, Antisocial Personality Disorder, and alcoholism in parents (Rutter & Giller, 1983; West & Prinz, 1987) and grandparents (Glueck &Glueck, 1968) also placed children at risk for behavior associated with conduct disorder.

Several features related to the interaction of parents with their children were identified as risk factors for conduct disorder. Extreme punishment practices (Kazdin, 1985); lax, erratic, and inconsistent discipline practices (McCord, McCord, &Zola, 1959); and poor supervision (Wilson, 1980) were implicated in

conduct disorder. Parents of antisocial children were more likely to give commands to their children, reward deviant behavior directly through attention and compliance, and ignore or provide aversive consequences for prosocial behavior (Patterson, Reid, & Dishion, 1992). Parents of antisocial youths reported less attachment and showed less warmth, affection, and emotional support than parents of nonreferred youths (Henggeler, 1989).

Peer factors associated with conduct disorder include peer rejection of children who exhibit antisocial behavior (Dodge, Cole, & Brakke, 1982; Ladd, Hart, & Price, 1990; Patterson, DeBaryshe, & Ramsey, 1989). Membership in deviant peer groups that become involved in serious delinquent and antisocial behavior was associated with peer rejection (Cairns, Cairns, Neckerman, Gest, & Gariepy, 1988; Dishion, Patterson, & Stoolmiller, 1991).

Not all individuals at risk for conduct disorder develop this condition. Their resilience may be attributed to protective factors that may attenuate or cancel the influence of risk factors (Kazdin, 1986). In many cases, protective factors (e.g., easy temperment, above average intelligence, academic success, good relationship with parents) represent the opposite of risk factors. Kazdin identified a good relationship with an emotionally responsive, caregiving adult as an important factor that often reduces risk for conduct disorder.

From a diagnostic perspective, DSM-IV identified psychopathological conditions for which youth with conduct disorder are at risk. The conditions were: Antisocial Personality Disorder, Mood or Anxiety Disorders, Somatoform Disorders, and Substance-Related Disorders.

Youth with Conduct Disorder and Reward Conditions

Theoretical and empirical work presented by Gray (1982, 1987) was reported to have been very influential in developing a line of research exploring the relationship between emotions, behavior, and performance within the context of instrumental learning paradigms (Shapiro & Hynd, 1993). Based on a series of experimental findings, Gray (1987) emphasized the utility of viewing personality dimensions in the terms of individual differences in conditionability and sensitivity to punishment versus reward. This led to the proposal that behavior is regulated primarily by two different but interacting conceptual brain systems: the behavioral inhibition system (BIS) and the behavioral activation (or reward) system (BAS). The two systems have different neurophysiological substrates and are responsive to different types of stimuli. The BIS is thought to inhibit behavior in response to cues of punishment, frustrative nonreward, or novelty, while the BAS is thought to activate behavior in response to cues of reward or nonpunishment.

Quay (1988a, 1988b, 1988c) proposed that conduct disorder involves a persistently overactive reward system (BAS) that predominates over the BIS and results in persistent reward-dominate behavior and difficulty in delaying gratification. Studies in which participants with conduct disorder responded strongly to rewards on instrumental learning tasks have supported Quay's proposal (Daugherty & Quay, 1991; Daughterty, Quay, &Ramos, 1991; Fonseca & Yule, 1995; Kalantari, Yule, & Gardner, 1990; Shapiro, Quay, Hogan, & Schwartz, 1988).

Shapiro et al. (1988) studied the reward-seeking behavior of children and adolescents with conduct disorder on either a card-playing task or a differential reinforcement for low-rate responding task. The participants were selected from a group of 56 public school students (age range 7 years, 6 months – 18 years) who were classified as seriously emotionally disturbed (SED). The participants were from predominantly lower-middle-class families based on sociodemographic information obtained from the participants' school records. Participants whose Full Scale IQs were below 70 were excluded from the study. A conduct disordered group of nine participant (5 male, 4 female; 5 White, 4 Black) was identified by selecting participants whose Conduct Disorder Scale T score (SED norms) on the Revised Behavioral Problem Checklist (RBPC; Quay & Peterson,

1987) was both 55 and was their most elevated scale. A comparison group of 10 participant (9 male, 1 female; 8 White, 2 Black) was identified by selecting participants whose RBPC Conduct Disorder Scale T score (SED norms) was 50 or below and whose T score on all of the other five RBPC was below 55. Both groups were equivalent with regard to level of intellectual functioning and age. Measures of the participants' response perseveration on a card playing task in which the probability of losing increased in a controlled manner (number of cards played) and delayed responding on a differential reinforcement for low-rate responding task, programmed with a 6 second delay, (efficiency ratio for the last trial block) were obtained. Participants were randomly assigned to receive either of the two tasks first. Money (a dime for winning on the card playing task and a nickel for the necessary delay on the responding task) was chosen as the reinforcer. The conduct disorder group performed more poorly on both tasks than the comparison group. The results were significant and supported Quay's reward dominance hypothesis of conduct disorder.

In a similar vein, Daugherty and Quay (1991) studied children's reward-seeking behavior on a correct door-opening task (a modified version of the card-playing task used by Shapiro et al. and considered more appropriate for younger participants) and a differential reinforcement for rewarding low-rate responding

task. The participants were selected from a nonclinical population of 814 elementary school students (age range 8.1 years – 13.8 years). The ethnic composition of all participants was 66% Hispanic, 26% Caucasian, and 8% Black. The RBPC was used to identify five groups of participants who exhibited behaviors consistent with conduct disorder, conduct disorder with attention-deficit/hyperactivity disorder, attention deficit disorder, anxiety-withdrawal disorder, and minimum behavior problems (the normal group). The groups were equivalent with regard to age, gender, and academic achievement. The participants earned or lost chips based on correct or incorrect responses on the door-opening task and earned chips on the delayed responding task. The chips were exchanged for a small prize after both tasks were completed. Both the conduct disorder group and the conduct disorder with attention-deficit /hyperactivity disorder group opened significantly more doors than did the normal group and tended to respond persistently despite worsening odds. No significant differences in efficiency ratio were found among the four clinical groups and the normal control group on the delay responding task. The investigators hypothesized that the failure to replicate earlier efficiency ratio differences may have been associated with the nonclinical participant population used in this

study. The results from the door-opening task supported Quay's reward dominance hypothesis of conduct disorder.

Shapiro and Hynd (1993) cited additional studies related to the reward dominance hypothesis. Kalantari, Yule, and Gardner (1990) extended the findings of perservative responding to a sample of preschool children in Iran. The 6 participants rated by their teachers as having persistent behavior problems played significantly more cards than the 6 participants in the normal control group. Daugherty, Quay, and Ramos (1991) found a strong tendency for their RBPC-defined conduct disorder group to respond perseveratively on the door-opening task in comparison to the conduct disorder with attention-deficit /hyperactivity disorder, attention deficit disorder, anxiety-withdrawal disorder, and normal control groups that did not differ significantly.

Fonseca and Yule (1995) conducted two studies that supported the hypothesis that children with conduct disorder are more sensitive to rewards than normal children. The participants in the first study consisted of 44 male juvenile delinquents (age range 12 years – 15 years) who qualified for a diagnosis of conduct disorder. This group was divided into equal size, aggressive and nonaggressive subgroups. A control group of 20 males (age range 11 years – 15 years) with no histories of antisocial behavior or police or court contacts were

selected from two youth clubs and one scout group. All participants were selected from the same inner city area. Both groups were equivalent in IQ. On a response perseveration card game task similar to the task used by Shapiro et al. (1988), the aggressive and nonaggressive groups played significantly more cards and earned significantly less money (used for rewards) than the normal group.

In the second study, the conduct disorder group consisted of 27 males (age range 7 years – 11 years) who either attended a special school for emotionally disturbed children or were treated in an outpatient or inpatient unit at a children's psychiatric hospital. All participants scored above the cutoff point for antisocial behavior on the Rutter Questionnaire for Teachers (Rutter, 1987) and met criteria for a diagnosis of conduct disorder on DSM-III-R and ICD-9. The control group consisted of 26 males (age range 7 years – 11 years) chosen from a school in the same area as the school for emotionally disturbed children. Control participants experienced no difficulties in adapting to school and scored below the cutoff score on the Rutter Questionnaire. The conduct disorder group had significantly lower IQs (M=95.0, SD=13.6) than the control group (M=107.4, SD=13.4). On the response perseveration card game task used in the first study, the conduct disorder group, like the aggressive and nonaggressive delinquent groups, played significantly more cards and earned significantly fewer rewards (candies instead

of money) than the control group. Both studies provided considerable support for Quay's reward dominance hypothesis of conduct disorder.

The number of studies involving the differential responsiveness to social rewards in which children and adolescents were used as participants was limited (Quay, 1986). Levin and Simmons (1962a) reported that a group of 15 severely emotionally disturbed, hyperactive, and aggressive boys in a residential setting were less responsive to social reinforcers than normal participants. In a study involving six participants from their first study, Levin and Simmons (1962b) compared the performance of participants who were reinforced by food alone with participants who were reinforced by food plus praise. Their results suggested that praise suppressed the response rate. It should be noted that the participants who received food performed significantly better than they had in the earlier study when praise was the sole reinforcer. The confidence in these results was limited by the small sample size.

Stewart (1972) studied the effects of social reinforcement on a sentence-building task presented to 78 institutionalized male delinquents (age range 16 years – 22 years). The participants were equally divided into three groups, undersocialized, socialized, or anxious-withdrawn, based on three psychometric measures (Peterson, Quay, & Tiffany, 1961; Quay, 1964a; Quay, 1964b) and

randomly assigned to either a frustration condition (working on an insoluble task) or a no-frustration condition (working on Porteus Maze [1985] tasks) that preceded the sentence-building tasks. Under the frustration condition, the anxious-withdrawn and socialized groups significantly increased, while the undersocialized group significantly decreased the use of reinforced verbs. When the conditioning was preceded by the no-frustration condition, the anxious-withdrawn and socialized groups did not significantly increase the use of reinforced verbs. In this situation, the undersocialized group continued to significantly decrease the use of reinforced verbs.

Dietrich (1976) studied the efforts of variable task structure (neutral and emotional content, simplicity, and complexity) and reward conditions (material and verbal) on the performance of 36 institutionalized delinquent boys (age range 10 years – 16 years, IQ range 71-133) on a verbal incentive paired-associate task. The participants were equally divided into three behavior classification groups (conduct disorder, personality disorder, and immature-inadequate) based on teacher ratings on the Behavior Problem Checklist (Quay, 1964b). As expected, the conduct disorder group preformed significantly better on the simple, neutral task under conditions of candy reward. Under verbal reinforcement, the same group preformed significantly better on the complex, neutral task. Dietrech

offered explanatory theories based on consolidation theory, novel stimulation, and practice effects for the unexpected beneficial effects of verbal praise.

Quay (1988) reported that the literature supported no clear general relationship between undersocialized conduct disorder and lowered responsiveness to social reward in children and youth. He added that this finding was consistent with the contradictory results from studies that investigated the sensitivity of adult psychopaths to verbal or social reinforcement.

Assessment

Identification of conduct disorder in children and adolescents relies on assessment to determine the presence of symptoms and characteristics associated with this condition. A variety of modalities and types of measures to identify conduct disorder were discussed by Kazdin (1986). The strengths, methodological weaknesses, and sources of bias for self-report measures, reports of significant others, peer evaluations, direct observation, and institutional and societal records were reported. It was recommended that in most instances, multiple modalities should be included in an assessment battery. Although children and adolescents were reported to rarely identify themselves as having a problem or needing treatment, they were reported to be able to provide information on their antisocial behavior.

Efforts to identify characteristic differences for youth with conduct disorders on individually administered intelligence test in comparison to SES-comparable groups with and without other disorders have not been successful (Quay, 1986). Similarly, efforts to show significant discrepancies between verbal and performance scores were not successful when adequate controls were utilized. In studies that established greater diagnostic specificity (e.g., McGee, Williams, & Silva, 1984; Richman & Lindgrin, 1981), hyperactive and aggressive 7-year-old participants and undersocialized youth with conduct disorder exhibited significant verbal weaknesses. Moffitt (1993b) suggested that the robust IQ deficit (one-half standard deviation) found in the literature between juvenile delinquents and their nondelinquent peers may be relevant to youth with conduct disorder. A study (Moffitt, 1990) reporting a 17-point mean deficit for externalizing youths with a long history of aggressive behavior compared to nondelinquent youths supported this inference.

Summary

Children and adolescents who meet DSM-IV (American Psychiatric Association, 1994) criteria for Conduct Disorder represent a group of emotionally disturbed, behavior disordered youth who exhibit a pattern of antisocial behaviors that occurs chronically and impairs the functioning of the individual (Kazdin,

1996). Specifically, the antisocial behaviors include aggressive acts, theft, vandalism, setting fires, lying, truancy, and running away. Youths who exhibit these problems comprise from one third to one half of all children and adolescent clinic referrals (Kazdin, Siegal, & Bass, 1990; Robins, 1981). The Institute of Medicine (1989) estimated that the rate of conduct disorder among children from ages 4 to 18 ranged from 2 to 6%.

Children diagnosed with conduct disorder can vary widely in terms of specific symptoms and associated features of the disorder (Kazdin, 1996). A number of risk factors increases the likelihood of the occurrence of the condition (Kazdin, 1996). Major categories of risk factors for conduct disorder include factors related to the child, the parent and family, the school, and the peers (Kazdin, 1996; Quay, 1986; Robins, 1991; Short & Shapiro, 1993).

Quay (1988a, 1988b, 1988c) proposed that conduct disorder involves a persistently overactive reward system that predominates over an underactive behavioral inhibition system and results in persistent reward-dominant behavior and difficulty in delaying gratification. Studies in which participants with conduct disorder respond strongly to rewards on instrumental learning tasks have supported Quay's proposal (Daugherty & Quay, 1991; Daugherty, Quay, & Ramos, 1991; Shapiro, Quay, Hogan, & Schwartz, 1988). In addition, test

incentives have been shown to improve the cognitive test performance of emotionally disturbed, behavior disordered youth (Jackson et al., 1979; Mayfield 1977; Saigh, 1982).

A variety of modalities and types of measures to identify conduct disorder were discussed by Kazdin (1996) including self-report measures, reports of significant others, peer evaluations, direct observation, and institutional and societal records. It was recommended that in most instances, multiple modalities should be included in an assessment battery.

CHAPTER 2

METHODOLOGY

Statement of the Problem

An implicit aim of the practice of school psychology involves obtaining the best estimates of examinee performance during standardized assessments (National Association of School Psychologists, 1992). Verbal praise and material rewards have been shown to be effective in improving the performance of youths with mild mental retardation during cognitive testing (Saigh, 1981a; Saigh & Payne, 1976, 1979; Terrell, Terrell, & Taylor, 1980). Although the literature suggests that conduct disorder youths are more responsive to material rewards (Shapiro & Hynd, 1993), only one study (Saigh, 1982) investigating the differential effects of verbal and material rewards on their cognitive performance was identified.

Rationale and Hypotheses

This study attempted to show that conduct disorder youths perform significantly better on an individually administered intelligence test when given material rewards in comparison to groups who receive verbal praise or neutral (i.e., non-evaluative) comments. Material rewards are expected to act as incentives to improve the level of motivation of youth with conduct disorders who

generally do not perform in an optimal manner on tests of cognitive ability. There is evidence to suggest that youths with conduct disorder respond to material rewards in a greater and more consistent manner than they respond to neutral comments on cognitive test performance (Saigh, 1982). Saigh also suggested that material incentives may be "a viable means of facilitating the WISC-R performance of adolescents with conduct disorder" (p.772)

With these points in mind, the following hypotheses regarding the participants' WISC-III performance were tested:

H01: The mean Full Scale IQ score of the material reward group will be significantly greater than mean Full Scale IQ scores of the verbal praise and the control groups.

H02: The mean Verbal Scale IQ score of the material reward group will be significantly greater than mean Verbal Scale IQ scores of the verbal praise and the control groups.

H03: The mean Performance Scale IQ score of the material reward group will be significantly greater than mean Performance Scale IQ scores of the verbal praise and the control groups.

H04: The mean Picture Completion scaled score of the material reward group will be significantly greater than mean Picture Completion scaled scores of the verbal praise and the control groups.

H05: The mean Information scaled score of the material reward group will be significantly greater than mean Information scaled scores of the verbal praise and the control groups.

H06: The mean Coding scaled score of the material reward group will be significantly greater than mean Coding scaled scores of the verbal praise and the control groups.

H07: The mean Similarities scaled score of the material reward group will be significantly greater than mean Similarities scaled scores of the verbal praise and the control groups.

H08: The mean Picture Arrangement scaled score of the material reward group will be significantly greater than mean Picture Arrangement scaled scores of the verbal praise and the control groups.

H09: The mean Arithmetic scaled score of the material reward group will be significantly greater than the mean Arithmetic scaled scores of the verbal praise and the control groups.

H10: The mean Block Design scaled score of the material reward group will be significantly greater than mean Block Design scaled scores of the verbal praise and the control groups.

H11: The mean Vocabulary scaled score of the material reward group will be significantly greater than mean Vocabulary scaled scores of the verbal praise and the control groups.

H12: The mean Object Assembly scaled score of the material reward group will be significantly greater than mean Object Assembly scaled scores of the verbal praise and the control groups.

H13: The mean Comprehension scaled score of the material reward group will be significantly greater than mean Comprehension scaled scores of the verbal praise and the control groups.

H14: The mean Full Scale IQ score of the verbal praise group will not be significantly different than the mean Full Scale IQ score of the control group.

H15: The mean Verbal Scale IQ score of the verbal praise group will not be significantly different than the mean Verbal Scale IQ score of the control group.

H16: The mean Performance Scale IQ score of the verbal praise group will not be significantly different than the mean Performance Scale IQ score of the control group.

H17: The mean Picture Completion scaled score of the verbal praise group will not be significantly different than the mean Picture Completion scaled score of the control group.

H18: The mean Information scaled score of the verbal praise group will not be significantly different than the mean Information scaled score of the control group.

H19: The mean Coding scaled score of the verbal praise group will not be significantly different than the mean Coding scaled score of the control group.

H20: The mean Similarities scaled score of the verbal praise group will not be significantly different than the mean Similarities scaled score of the control group.

H21: The mean Picture Arrangement scaled score of the verbal praise group will not be significantly different than the mean Picture Arrangement scaled score of the control group.

H22: The mean Arithmetic scaled score of the verbal praise group will not be significantly different than the mean Arithmetic scaled score of the control group.

H23: The mean Block Design scaled score of the verbal praise group will not be significantly different than the mean Block Design scaled score of the control group.

H24: The mean Vocabulary scaled score of the verbal praise group will not be significantly different than the mean Vocabulary scaled score of the control group.

H25: The mean Object Assembly scaled score of the verbal praise group will not be significantly different than the mean Object Assembly scaled score of the control group.

H26: The mean Comprehension scaled score of the verbal praise group will not be significantly different than the mean Comprehension scaled score of the control group.

H27: The mean Post Hoc Follow-up Questionnaire total score of the material reward group will be significantly greater than mean Post Hoc Follow-up Questionnaire total scores of the verbal praise and the control groups.

H28: The mean Post Hoc Follow-up Questionnaire total score of the verbal praise group will not be significantly different than the mean Post Hoc Follow-up Questionnaire total score of the control group.

Method

Reward Setting and Participants

Participants were selected from two residential treatment centers located in Westchester County, New York. All participants attended schools that served the population of the respective residential treatment center. The schools, which are located on the campuses of the residential treatment centers, included day school programs. Day students who attended the schools were included in the sampling pool. The sampling pool of potential participants consisted of 636 youths (age range 8 years – 22 years, 406 males and 230 females).

At one residential treatment center, one of two high school programs served 141 students (age range 13 years- 22 years) who exhibited more aggressive and defiant behavior. The other high school program served 85 students (age range 13 years - 22 years) who were more emotionally fragile and who exhibited explosive and noncompliant behavior. The middle school served 60 students (age range 8 years - 12 years). These students exhibited defiant, noncompliant, and aggressive behaviors. The schools served 238 males and 48 females.

At the other residential treatment center, the high school program served 175 students (age range 13 years - 22 years), the junior high school served 95 students (age range 12 years - 14 years), and the elementary school served 80 students (age

range 8 years - 12 years). Most students exhibited aggressive, noncompliant, and defiant behavior. The schools served 168 males and 182 females

Most of the students were classified as Seriously Emotionally Disturbed Special Education students based on New York State Special Education Rules and Regulations. The majority of these students experienced learning difficulties in addition to emotional problems.

Residents and students were placed in the programs by social services, the courts, and/or local school districts. Almost all of the participants experienced chaotic and/or abusive home situations. The majority of participants (98%) came from urban areas. Sixty-six percent of the students were Black; 8% were White; 22% were Hispanic; 1% was Asian; and 3% were of mixed descent.

Diagnostic Measures

Diagnostic Interview for Children and Adolescents: Child Version and Adolescent Version, DSM-III-R and DSM-IV Editions (DICA, Reich, Leacock, & Shanfeld, 1995)

The DICA is a structured clinical interview that was utilized to diagnose a number of related disorders that may impact on children's intellectual performance. The DICA is divided into a series of modules that are indicative of specific childhood disorders, according to DSM-IV diagnostic criteria. Selected

subtests from the DICA will be used to determine the presence or absence of Attention-Deficit/Hyperactivity Disorder, Conduct Disorder, Mood Disorders, Psychotic Disorders, and Substance Dependence among participants. Reliability and validity estimates of the DICA suggest that the structured format provides a sound diagnostic tool, far superior to informal interview methods. Inter-rater reliability estimates on the DICA indicated 85% agreement among 10 independent examiners exposed to videotaped interviews (Herjanic & Reich, 1982). Test-retest reliability estimates resulted in 89% agreement for individual symptoms, over several months. Validity for the DICA was supported by discriminant validity studies (kappa = .76), and by comparison of clinically derived discharge diagnoses of inpatients conducted by psychiatric personnel. One author (W. Reich) reported moderate kappa coefficients for the DICA-R (personal communication, July, 1997). Data collected from 48 subjects between the ages of 13 and 17 resulted in kappa values of .56, .78, and .82 for ADHD, ODD, and CD modules respectively.

Clinical Interview

Detailed background information was collected for each experimental participant through informal interviews with participants and clinical staff and through careful review of the participant's educational/residential records. The

investigator determined the participant's age, school, residential status, gender, grade, and ethnicity. The clinical interview also included the parent/guardian's marital status, occupation, level of education, and geographic location of residence (i.e., urban, suburban, rural) and the family composition. In addition, the interview screened for prior head trauma, seizure disorders, epilepsy, and current medications. The record reviews and clinical interviews were also used to screen for potential confounding conditions such as dementia, delirium, mental retardation, and pervasive developmental disorder, and to corroborate diagnostic conditions assessed on the DICA. The background information was transferred to a standardized Research Cover Sheet (See Appendix A).

Material Reward Menu

The Material Reward Menu was used to identify meaningful rewards for participants. A list of rewards ranging in value from $1.00 to $10.00 used as reinforcers with the participants in their school and residential programs was developed by inquiring among clinicians and reviewing existing behavior management programs (See Appendix B). During the clinical interview, the participants were instructed to indicate the five rewards (each worth approximately $10.00) that they preferred the most as described in the Material Reward Menu. Rewards included gift certificates for food, batteries, cassette

tapes, toiletry items, and t-shirts. Space was provided for the participant to identify rewards that were not listed on the menu. The availability and appropriateness of these rewards were discussed, and items that could not be included on the menu were identified.

Dependent Variables

Wechsler Intelligence Scale for Children – Third Edition (WISC-III, Wechsler, 1991)

Intellectual performance was measured by the WISC-III, an individually administered clinical instrument for assessing the intellectual ability of school-age children. The WISC-III represents the most widely utilized and accepted intellectual measure currently in use in the United States (Sattler, 1992). The WISC-III is organized into a four factor model (Wechsler, 1991) which includes Factor I: Verbal Comprehension (VC); Factor II: Perceptual Organization (PO); Factor III: Freedom from Distractibility (FD); and Factor IV: Processing Speed (PS). Figure 1 presents the WISC-III subtest composition of each factor-based index:

Figure 1

Indices Derived from Factor Analysis of Subtests

Subtest	VC	PO	FD	PS
Picture Completion		X		
Information	X			
Coding				X
Similarities	X			
Picture Arrangement		X		
Arithmetic			X	
Block Design		X		
Vocabulary	X			
Object Assembly		X		
Comprehension	X			
Symbol Search				X
Digit Span			X	

The WISC-III contains 13 subtests, six of which form the Verbal Scale (Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit

Span) and seven of which form the Performance Scale (Picture Completion, Coding, Picture Arrangement, Block Design, Object Assembly, Symbol Search, and Mazes). Performances on the 13 subtests are tabulated into the Verbal IQ (VIQ), Performance IQ (PIQ), and Full Scale IQ (FSIQ) Scores. In addition to the three IQ scores around which the WISC-III is organized, factor-based index scores are also calculated (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed).

The WISC-III was standardized with a sample of 2,200 cases, including 200 children from each of eleven age groups ranging from six to sixteen years (Wechsler, 1991). The sample included 100 males and 100 females for each age group, and was proportionate for such factors as race, parental education, and geographic region, according to U.S. census data. Sattler (1992) notes that the WISC-III has excellent reliability. The three scales have internal consistency reliability coefficients of .89 or above across the entire standardization group. Split-half correlations for the individual subtests range from .70 to .87. Average internal consistency for the Full Scale IQ Score is .96. Test-retest reliability was assessed with 353 children from six age groups, with intervals ranging from 12 to 63 days (Wechsler, 1991). Stability coefficients for the Full Scale IQ ranged from .92 to .95. Inter-scorer agreement on all the WISC-III subtests averaged in the

high .90s. Construct validity, as determined by factor analysis, has determined that the Full Scale IQ Score represents a fair estimate of g, or global ability. Concurrent and predictive validity studies reported in Wechsler (1991) found the WISC-III to be highly correlated with other intellectual measures. Concurrent validity estimates indicate the WISC-III FSIQ to be highly correlated (r = .89) with the earlier WISC-R FSIQ.

A detailed description of each WISC-III subtest, IQ, and index scores is provided below:

Information.

The Information subtest contains 30 questions that sample a broad range of general knowledge, including names of objects, dates, and historical, scientific, and geographical facts. Performance on this subtest is a good measure of general ability and is dependent on subjects' long-term memory, social and cultural background, and the extent of their formal and informal educational opportunity. Weschsler (1991) notes the Information subtest has a split-half reliability coefficient of .84, and correlates highly with the Vocabulary (r = .70) and Similarities (r = .66) subtests, and the VIQ (r = .75).

Similarities.

The Similarities subtest contains 19 pairs of words; the child is required to explain the similarity between two items in each pair. After several initial one point items, the subject may earn either zero, one, or two points, depending on the conceptual level of his or her response. This subtest most closely measures verbal concept formation, although Similarities may also tap long-term memory, expressive language skills, and cultural opportunities. Similarities has a split-half reliability coefficient of .81 and correlates most highly with Vocabulary (r = .69) and Information (r = .66) (Wechsler, 1991). It correlates moderately with the FSIQ (r = .72) and VIQ (r = .75), and to a lesser degree with the PIQ (r = .55).

Arithmetic.

The Arithmetic subtest contains 24 mathematics problems, each of which must be solved without paper and pencil. Initial problems may involve simple counting or operations (addition, subtraction), while the later problems may involve multi-step word problems. Each item contains a time limit, with bonus points available on the final six items for swift problem completion. The Arithmetic subtest requires the child to follow verbal directions, concentrate on selected parts of the word problem, use numerical problem solving, and work within a time limit. The subtest is a fair measure of general ability, and has a high loading for the Freedom

from Distractibility factor. Arithmrtic has a split-half reliability coefficient of .78, and correlates most highly with Similarities (r = .55) and Information (r = .57) (Wechsler, 1991). It correlates moderately with the FSIQ (r = .65) and VIQ (r = .62), and to a lesser degree with the PIQ (r = .54).

Vocabulary.

The Vocabulary subtest contains 30 words arranged in order of increasing difficulty. The examinee is asked to explain orally (define) each word presented, and the response is awarded zero, one, or two points, depending on the accuracy of the response. This test of word knowledge may tap a variety of factors, such as verbal fluency, learning ability, richness of ideas, long-term memory, and concept formation. The Vocabulary subtest is the best measure of general ability and is the most reliable WISC-III subtest. Vocabulary has a split-half reliability coefficient of .87, and correlates most highly with Similarities (r = .69), Information (r = .70), and Comprehension (r = .64) (Wechsler, 1991). It correlates moderately with the FSIQ (r = .74) and VIQ (r = .78), and to a lesser degree with the PIQ (r = .56).

Comprehension.

The Comprehension subtest contains 18 questions that deal with problem situations involving social and ethical mores, interpersonal relations, knowledge

of one's body, and an awareness of the world around us. Comprehension involves understanding a given situation and utilizing proper judgement and problem solving skills. Responses may reflect a child's moral and cultural knowledge, social judgement, common sense, and the examinee's grasp of social conventionality. The Comprehension subtest is good measure of general ability and is highly loaded on the Verbal Comprehension factor. Comprehension has a split-half reliability coefficient of .77 and correlates most highly with Vocabulary ($\underline{r}$ = .64), Similarities ($\underline{r}$ = .59), and Information ($\underline{r}$ = .56) (Wechsler, 1991). It correlates moderately with the FSIQ ($\underline{r}$ = .64) and VIQ ($\underline{r}$ = .67), and to a lesser degree with the PIQ ($\underline{r}$ = .49).

Digit Span.

For the supplementary Digit Span subtest, the child listens to a series of digits given orally by the examiner and then is asked to repeat the digits from memory. In the first part, Digits Forward, the examinees are asked to repeat strings of numbers from two to nine digits in length. Digits Backwards, which follows, involves recalling series of numbers from two to eight digits in length, in reverse order. Digit Span is a reliable measure of short-term auditory memory and concentration. While Digits Forward primarily involves rote learning and memory, Digits Backwards involves more advanced mental imagery and

numerical sequencing. Digit Span has a split-half reliability coefficient of .85 and correlates most highly with Arithmetic (r =.43) (Wechsler, 1991). It correlates poorly with the FSIQ (r = .43), VIQ (r = .42), and PIQ (r = .35).

Picture Completion.

The Picture Completion subtest contains 30 drawings of objects from everyday life, each missing an important element which the child must identify. The examinee must identify and name the missing object within a twenty-second time limit. This subtest measures visual discrimination, attention to detail, reasoning, alertness, and concentration. Also, Picture Completion may measure perceptual and conceptual abilities and the ability to discriminate essential from non-essential details. Picture Completion is a fair measure of general ability and draws highly from the Perceptual Organization factor. Picture Completion has a split-half reliability coefficient of .77 and correlates most highly with Block Design (r = .52), Object Assembly (r = .49), and Information (r = .47) (Wechsler, 1991). It correlates poorly with the FSIQ (r = .58), VIQ (r = .52), and PIQ (r = .54).

Coding.

The Coding subtest requires the child to copy symbols that are paired with other symbols, such as numerals. Coding A is administered to children under

eight years of age, while the more complex Coding B is utilized for examinees 8 to 16 years of age. Coding measures the examinee's ability to learn an unfamiliar task and taps the child's clerical speed and accuracy, visual-motor coordination, short-term memory, and concentration. Coding also measures the child's performance on speeded (timed) tasks. Coding is a poor measure of general ability, but does load on the Freedom from Distractibility factor. Coding has a split-half reliability coefficient of .79 and correlates most highly with Symbol Search (r = .53) (Wechsler, 1991). It has low correlation with the FSIQ (r = .33), VIQ (r =.29), and PIQ (r = .32).

Picture Arrangement.

For Picture Arrangement, the examinee is required to place a series of related pictures into a logical sequence, within a specified period of time. The 14 series of pictures are similar to short comic strips, but without words or written cues. This subtest measures the child's ability to comprehend and evaluate a total situation, while using trial-and-error experimentation to solve the problem. Picture Arrangement measures the examinee's ability to form a whole from its parts, planning ability, visual organization, and reasoning. This subtest represents only a fair measure of general ability, but is loaded on the Perceptual Organization factor. Picture Arrangement has a split-half reliability coefficient of

.76 and correlates most highly with Block Design (r = .41), Information (r = .40), Vocabulary (r =.40), and Similarities (r = .39) (Wechsler, 1991). It has a low correlation with the FSIQ (r =.52), VIQ (r = .45), and PIQ (r =.49).

Block Design.

The Block Design subtest contains12 items and consists of two-dimensional red and white printed figures and three-dimensional blocks used to form the pictured figures. This subtest measures the child's ability to perceive and analyze forms by breaking down the whole design into component parts and then assembling the blocks to represent the whole figure. Block Design is the best measure of general ability among the Performance Scale subtests and is highly related to the Perceptual Organization factor. Block Design has a split-half reliability coefficient of .87 and correlates most highly with Object Assembly (r = .61), Picture Completion (r = .52), and Arithmetic (r = .52) (Wechsler, 1991). It has a moderate correlation with the FSIQ (r = .66), and the PIQ (r = .65), and to a lesser degree with VIQ (r = .57).

Object Assembly.

The Object Assembly subtest requires that the examinee assemble jigsaw puzzle pieces into common objects, within a specific time limit. Additional points

are awarded for swift completion of the puzzles. Object Assembly tests the child's ability to synthesize objects into a whole, as well as the examinee's visual-motor dexterity and perceptual ability. This subtest represents a fair estimate of general ability and is loaded on the Perceptual Organization factor. Object Assembly has a split-half reliability coefficient of .69 and correlates most highly with Block Design (r = .61) (Wechsler, 1991). It has a somewhat low correlation with the FSIQ (r = .58), a moderate correlation with the PIQ (r = .60), and a low correlation with the VIQ (r = .48).

Symbol Search.

The supplementary Symbol Search subtest requires the examinee to quickly recognize familiar symbols present within an array of symbols, within a limited amount of time. The child is asked to identify (Yes/No) whether a simple geometric figure appears embedded within a series of similar figures. Part A is given to children under 8 years, while the more complex Part B is administered to children 8 to 16years of age. This subtest measures attention to detail, clerical speed, and concentration. Symbol Search is not strongly related to general ability, but is loaded on the Processing Speed factor. Symbol Search has a split-half reliability coefficient of .76 and correlates most highly with Coding (r = .53) (Wechsler, 1991). It has a somewhat low correlation with the FSIQ

(r = .56), PIQ (r = .58), and VIQ (r = .44).

Mazes.

The supplementary Mazes subtest consists of ten maze problems where the child is asked to draw a pencil line from the center of the maze to the outside, with few clerical errors and within a limited amount of time. In order to complete each maze successfully, the examinee must follow directions, use planning and strategy, and demonstrate good fine-motor dexterity. Mazes is a poor indicator of general ability, but does contribute moderately to the Perceptual Organization factor. According to Wechsler (1991), mazes has a split-half reliability coefficient of .70 and correlates most highly with Block Design (r = .31). It has a low correlation with the FSIQ (r = .31), PIQ (r = .35), and VIQ (r = .23).

Verbal Comprehension Index.

This index score is formed by combining the results of the Information, Similarities, Vocabulary, and Comprehension subtest scores of the WISC-III. Verbal Comprehension measures verbal knowledge and understanding obtained by formal and informal education, and reflects the application of verbal skills to new situations. This measure of verbal ability is a variable common to most of

the Verbal Scale subtests. According to Wechsler (1991), the split-half reliability coefficient for this index is .94 with a stability coefficient of .93.

Perceptual Organization Index.

This index score is formed by combining the results of the Picture Completion, Picture Arrangement, Block Design, and Object Assembly subtests. This index score measures the examinee's ability to interpret and organize visually perceived material while working under a time limit. The index score measures a factor common in most of the Performance Scale subtests. According to Wechsler (1991), the split-half reliability coefficient for this index is .90, with a stability coefficient of .87.

Freedom from Distractibility Index.

This index score is formed by combining the results of the Arithmetic and supplementary Digit Span subtests. The Freedom from Distractibility score measures the examinee's ability to attend and concentrate, but also may be related to numerical proficiency and sequencing skills. According to Wechsler (1991), the split-half reliability coefficient for this index is .87, with a stability coefficient of .82.

Processing Speed Index.

This index score is formed by combining the results of the Coding and supplementary Symbol Search subtests. Processing Speed involves the examinee's ability to employ a high degree of attention and concentration in processing information rapidly by scanning an array. According to Wechsler (1991), the split-half reliability coefficient for this index is .85, with a stability coefficient of .84.

Verbal IQ.

The Verbal IQ is formed by combining the results of the Verbal Scale subtests. The deviation IQ is determined by comparing the examinee's score with those earned by a representative sample of age peers. The Verbal IQ measures the examinee's performance across the full battery of verbal and verbal comprehension tasks. According to Wechsler (1991), the split-half reliability coefficient for the Verbal IQ is .95, with a stability coefficient of .94.

Performance IQ.

The Performance IQ is formed by combining the results of the Performance Scale subtests. The deviation IQ is determined by comparing the examinee's score with those earned by a representative sample of age peers. The Performance IQ measures the examinee's performance across the full battery of

tasks involving performance, perceptual organization, and non-verbal abilities. According to Wechsler (1991), the split-half reliability coefficient for the Performance IQ is .91, with a stability coefficient of .87.

Full Scale IQ.

The Full Scale IQ is formed by combining the results of the Verbal and Performance subtests, excluding the supplementary subtests under normal test conditions. The deviation IQ is determined by comparing the examinee's score with those earned by a representative sample of age peers. The Full Scale IQ represents the most reliable and accurate predictor of general ability. This score is also highly related to scholastic aptitude and the child's potential for academic success. According to Wechsler (1991), the split-half reliability coefficient for the Full Scale IQ is .96, with a stability coefficient of .94.

Post Hoc Follow-up Questionnaire (PHFQ)

A 20-item questionnaire was developed by the investigator to provide information related to the thoughts and feelings experienced by the participants during testing (See Appendix C). The thoughts and feelings referenced in the items were identified in the literature as variables that can affect test performance and that are influenced by incentive conditions (Anastasi, 1997; Goh & Lund, 1977; Graham, 1971; Pollock, 1990; Saigh & Payne, 1979; Saigh,1981a).

Participants marked each item on a 4-step Likert-type scale corresponding to "strongly agree," "agree," "disagree," and "strongly disagree". Responses were scored 4, 3, 2, or 1 indicating favorable to less favorable thoughts and feelings. On 14 items, agreement earned a high score; and on 6 items, disagreement earned a high score. The sum of the item credits represented the participant's total score (ranging from 20 to 80).

SES Index

Hollingshead Four Factor Index of Social Status (Hollingshead, 1975)

The Hollingshead Four Factor Index of Social Status (See Appendix D) was considered as a possible covariate, since a positive correlation ($r = .33$) was reported between a family's socioeconomic status and children's intelligence scores (Sattler, 1992). The Hollingshead Index is conceptually based on a multidimensional model of social status, derived from four factors: (1) occupation; (2) level of education; (3) marital status; and (4) gender. The Hollingshead Index provides a composite score (ranging from 8 to 66) derived from the first three factors. Parental occupation is graded on a 9-point scale ranging from manual laborers (level 1) to higher executives (level 9), based on United States census ratings. Education level, based on the number of years of school achievement, is scored on a 7-point scale ranging from less than seventh

grade (level 1) to graduate/professional training (level 7). The third factor, marital status, is used to determine whether one or two parent's occupational data should be utilized in calculating a SES score. For example, Hollingshead Index scores may be based on one family member's occupation (i.e., single, head of household) or an average of two occupations (i.e., two spouses gainfully employed). The last factor, gender, is not considered in the Hollingshead scoring.

The Hollingshead Index has been utilized in a variety of psychological and sociological studies to examine such factors as juvenile delinquency (Gray-Ray & Ray, 1990), maladaptive behavior (Peniston & McLean, 1979), behavioral disorders (Suh & Carlson, 1977), language functioning (Geffner & Hochberg, 1976), and visual-motor coordination (Amante, 1975). Validity data (construct) for both education and occupation is based upon 1970 census data and the National Opinion Research Council (NORC), and ranges from .835 (males) to .849 (females). The Hollingshead index and the NORC prestige index are correlated at $r = .927$. Gottfried (1985) reported that the Hollingshead index was extremely reliable in test-retest conditions, and was highly correlated ($r = .672$ - .781) with other SES measures. Gottfried (1986) concluded that the Hollingshead Index represents a "highly reliable and valid measure of socioeconomic differentiation in the United States" (p. 86).

Procedure

Examiners

The 6 examiners (3 male, 3 female) who administered the WISC-III were New York State certified school psychologists with extensive training and experience in administering that instrument. Because of time and availability constraints, the principal investigator served as an examiner (Examiner 2) for seven participants when other examiners were not available. The other examiners tested 5, 7, 11, 16, and 17 participants. All examiners (except for the principal investigator) were naive regarding the specific hypotheses of the study. The principal investigator provided the examiners with a Procedure Manual for the study (See Appendix E) that included specific procedures for each of the three testing conditions (See Appendix F). In addition, the principal investigator instructed the examiners on the test procedures, verbal statements, and schedule of reinforcement to be used with the participants in each condition. Modeling of the procedures was provided for the examiners by the investigator. The principal investigator observed the examiners practice the procedures until mastery was demonstrated by the examiners.

Participants were assigned to examiners for testing based on examiner availability at the time when a participant became eligible and available for the

study. If more than one examiner was available to work with a participant, the participant was randomly assigned to an examiner. This method of assigning participants to examiners resulted in Examiners 1, 2, 3, 4, 5, and 6 testing 7, 7, 16, 17, 5, and 11 participants respectively. A Chi-square test performed on the number of participants assigned to each examiner showed that the number of participants assigned to each examiner was not equally distributed (χ^2 [5, $\underline{N}$ = 63] = 12.143, $\underline{p}$ = .033).

Table 1 presents the means and standard deviations of the WISC-III Full Scale IQ scores of the male and female participants tested by each examiner.

Table 1

Means and Standard Deviations of the WISC-III Full Scale IQ Scores of the Male and Female Participants Tested by Each Examiner

Examiner	Male			Female			Total		
	n	M	SD	n	M	SD	n	M	SD
1	6	86.17	12.43	1	78.00	-	7	85.00	11.76
2	3	84.00	21.93	4	74.00	14.07	7	78.29	16.97
3	5	76.40	15.24	11	78.73	10.87	16	78.00	11.92
4	6	88.50	11.52	11	84.27	12.86	17	85.76	12.21
5	4	69.00	8.75	1	84.00	-	5	72.00	10.12
6	4	83.00	7.35	7	80.29	8.99	11	81.27	8.16
Total	28	81.79	13.51	35	80.37	11.24	63	81.00	12.21

A one-way ANOVA revealed no significant differences between the examiners by participant Full Scale IQ scores (F [5, 57] = 1.538, p = .193). Univariate ANOVA revealed no significant examiner differences between the examiners by gender for Full Scale IQ scores (F [5, 51] = .612, p = .691).

Table 2 presents the means and standard deviations of the Post Hoc Follow-up Questionnaire scores of the male and female participants tested by each examiner.

Table 2

Means and Standard Deviations of the Post Hoc Follow-up Questionnaire Scores of the Male and Female Participants Tested by Each Examiner

Examiner	Male			Female			Total		
	n	M	SD	n	M	SD	n	M	SD
1	6	59.00	4.38	1	63.00	-	7	59.57	4.28
2	3	57.00	6.00	4	62.25	8.81	7	60.00	7.66
3	5	63.80	12.93	11	57.27	12.15	16	59.31	12.36
4	6	64.17	8.18	11	65.18	8.08	17	64.82	7.87
5	4	63.00	8.91	1	54.00	-	5	61.20	8.71
6	4	60.00	4.97	7	59.86	7.97	11	59.91	6.74
Total	28	61.46	7.86	35	60.91	9.64	63	61.16	8.83

A one-way ANOVA revealed no significant differences between the examiners by participant Post Hoc Follow-up Questionnaire Total Scores (F [5, 57] = .828, p = .535). Univariate ANOVA revealed no significant examiner differences

between the examiners by gender for Post Hoc Follow-up Questionnaire Total Scores (F [5, 51] = .633, p = .676).

Informed Consent and Confidentiality

The parents or legal guardians of each participant received a cover letter and a consent form (See Appendices G and H) that explained the purpose of the investigation, procedures to be utilized, and potential drawbacks and benefits of the procedures. Both parental and participant assent (See Appendix I) were obtained in writing prior to data collection. Participants were informed that they were free to withdraw from the study at any time, without reprimand. All data was confidential and stored in a locked file by the investigator. Furthermore, after each test administration, the investigator transformed scores to research cover sheets that did not include the participants' names.

Participant Selection

Students between the ages of 12 through 16 years (n = 427) who attended schools at two residential/day treatment programs were invited to participate in the study. Based on a review of records, participants who were taking psychotropic medication (n = 160) or who were administered the WISC-III within the past 12 months (n = 41) were excluded from the study. This selection process reduced the sampling pool of potential participants from 636 to 226 youths who

were randomly assigned to one of three treatment groups (i.e., standard administration, verbal praise, or material reward) for the diagnostic interview and administration of the WISC-III. The sampling pool was further reduced to 86 participants because 140 participants were excluded from the diagnostic interview for the following reasons: parents or guardians did not respond to the request for their consent for approval (n = 73); discharge from the program (n = 24); absent without leave from the program (n = 22); refusal by the participant (n= 14); refusal by the parent or guardian (n = 4); and restriction by the participant's therapist for clinical reasons (n = 3).

Following the provisionary parental consent and participant assent, the conduct disorder module from the DICA was administered by the investigator to each student in order to identify participants who met DSM-IV criteria for Conduct Disorder. Information from the participant's record was used to corroborate the diagnosis. The DICA Attention-Deficit/Hyperactivity Disorder, Mood Disorders, Psychotic Disorders, and Substance Dependence modules, as well as clinical interviews and record reviews were used to exclude participants with comorbid Attention-Deficit/Hyperactivity Disorder, Mood Disorders, Psychotic Disorders ,Substance Dependence, Dementia, Delirium, Mental Retardation, or Pervasive Developmental Disorder. Similarly, participants who had a history of head

trauma, seizure disorders, or epilepsy, were excluded from the study. Exclusionary criteria was established to avoid the confounding effects that intellectual impairments associated with the following conditions may have on the results of this study: Attention-Deficit/Hyperactivity Disorder (Barclay, 1990), Mood Disorders (Slotkin, Forehand, Fauber, McCombs, & Long, 1988), Psychotic Disorders ,(Kendall & Hammen, 1996), Substance Dependence, (Kendall & Hammen, 1996), Mental Retardation, Pervasive Developmental Disorder, and head trauma (Hodges & Plow, 1990).

Twenty participants were excluded because they did not meet diagnostic criteria for Conduct Disorder ($\underline{n} = 8$) or because they were determined to be comorbid for Attention-Deficit/Hyperactivity Disorder ($\underline{n} = 6$) or a Mood Disorder ($\underline{n} = 6$). In addition, three participants who met diagnostic criteria for eligibility refused to complete the WISC-III.

Three groups of 21 adolescents with conduct disorder were obtained from the sample of 63 eligible participants who were randomly assigned to the standard administration, verbal praise, and material reward conditions. Cohn's (1988) power analysis procedure determined that 81 participants are needed to demonstrate adequate power (power = .80) and a large effect size with alpha = .05 for a three group comparison design with 10 dependent variables using a

multivariate analysis of variance procedure. Lower power (power = .50) was predicted for three groups of 21 participants who were eligible and available for the study.

Table 3 presents the distribution of the sample by race/ethnicity.

Table 3

Distribution of Sample by Race/Ethnicity

		Condition		
	Overall (n =63)	Standard (n = 21)	Verbal Praise (n = 21)	Material Reward (n = 21)
Black	44 (70%)	16 (76%)	11 (52%)	17 (81%)
Hispanic	18 (29%)	5 (24%)	9 (43%)	4 (19%)
Interracial	1 (1%)	0 (0%)	1 (5%)	0 (0%)

A Chi – square analysis revealed no significant difference between the conditions by race/ethnicity ($\chi^2 = 5.742$ (4) , p = .219). The mean WISC-III Full Scale IQ scores of race/ethnic groups (Black, M = 81.54, SD = 12.01; Hispanic, M = 79.00, SD = 12.88; Interracial, M = 93.00) were comparable based on one-way ANOVA (F [2,60] = .762, p = .471). The mean Post Hoc Follow-up Questionnaire scores of Blacks (M = 60.30, SD = 9.23), Hispanics (M = 62.72,

SD = 7.66), and Interracial participants (M = 71) were comparable (one-way ANOVA, F [2,60] = 1.119, p = .333).

The overall number of males in the sample (n = 28) did not differ significantly from the overall number of females (n = 35) based on a Binomial test for equal distribution (p = .45). Table 4 presents the distribution of the sample by gender.

Table 4

Distribution of Sample by Gender

	Condition			
	Overall (n = 63)	Standard (n = 21)	Verbal Praise (n = 21)	Material Reward (n = 21)
Male	28 (44%)	9 (43%)	8 (38%)	11 (52%)
Female	35 (56%)	12 (57%)	13 (62%)	10 (48%)

A Chi – square analysis revealed no significant difference between the conditions by gender (χ^2 = 900 (2), p = .638). The mean WISC-III Full Scale IQ scores of males (M = 81.76, SD = 13.51) was comparable to the mean WISC-III Full Scale IQ scores of females (M = 80.37, SD = 11.24), (t [61] = .454, p, 2-tailed = .652). The mean Post Hoc Follow-up Questionnaire scores of males (M = 61.46, SD = 7.86) was comparable to the mean Post Hoc Follow-up

Questionnaire scores of females (M = 60.91, SD = 9.64), (t [61] = .244, p, 2-tailed = .808).

The overall number of participants from Program A (n = 54) was significantly greater than the overall number of participants from Program B (n = 9) based on a Binomial test for equal distribution (p < .01). Table 5 presents the distribution of the sample by program.

Table 5

Distribution of Sample by Program

	Condition			
	Overall (n = 63)	Standard (n = 21)	Verbal Praise (n = 21)	Material Reward (n = 21)
Program A	54 (86%)	15 (71%)	19 (91%)	20 (95%)
Program B	9 (14%)	6 (29%)	2 (9%)	1 (5%)

The mean WISC-III Full Scale IQ scores of participants from Program A (M = 80.83, SD = 12.36) was comparable to the mean WISC-III Full Scale IQ scores of participants from Program B (M = 82.00, SD = 11.93), (t [61] = -.263, p, 2-tailed = .793). The mean Post Hoc Follow-up Questionnaire scores of participants from Program A (M = 61.41, SD = 9.40) was comparable to the mean Post Hoc

Follow-up Questionnaire scores of participants from Program B (M = 59.67, SD = 4.00), (t [26] = .942, p, 2-tailed = .355).

The mean age for participant from all conditions was 15 years, 5 months (range, 12 years, 6 months – 16 years, 11 months). The median age was 15 years, 9 months. Table 6 presents the mean ages and standard deviations of the sample (in months) by treatment condition.

Table 6

Mean Ages and Standard Deviations of the Sample by Treatment Condition

Condition	Mean (months)	SD (months)
Standard (n = 21)	184.67	14.78
Verbal Praise (n = 21)	185.10	14.91
Material Reward (n = 21)	185.90	11.13
Overall (n = 63)	185.22	13.50

One-way ANOVA revealed no significant differences between the conditions by participants' age (F [2,60] = .044, p = .957).

The investigator collected information from each participant's record regarding parental education and occupation. This information was used to calculate the

participants' Hollingshead Index ratings . The mean Hollingshead Index score for the sample group was 19.48 (SD = 10.42), with a minimum score of 8 and a maximum score of 56. Tables 7 presents the means and standard deviations of the sample for Hollingshead Index scores.

Table 7

Means and Standard Deviations for Hollingshead Index scores of the Sample by Treatment Condition

Condition	Mean	SD
Standard (n = 21)	21.10	11.28
Verbal Praise (n = 1)	20.52	11.97
Material Reward (n = 21)	16.81	7.37
Overall (n = 63)	19.48	10.42

One-way ANOVA revealed no significant differences between the conditions by Hollingshead Index scores (F [2,60] = 1.050, p = .356). A Pearson correlation between the Hollingshead Index scores and the WISC-III Full Scale IQ scores of the sample was calculated and showed a weak relationship between the measures (r = .176)

Reward Selection

During the interview to determine a participant's diagnostic eligibility, the investigator determined effective and appropriate reinforcers for each participant who indicated his or her preferences for reinforcers on the Material Reward Menu. Rewards that were medically or culturally contra-indicated for a participant were excluded. This procedure was completed during the clinical interview to avoid establishing the expectation of rewards for participants in the standard administration and verbal praise conditions.

Treatment Integrity

The examiners audiotaped their administration of the ten standard, non-supplementary WISC-III subtests for all participants for treatment integrity purposes. Two certified school psychologists, who were naive regarding the specific hypotheses of the study, independently rated the tape-recorded WISC-III administration for 15 randomly selected participants (5 from each experimental condition) for compliance with the experimental protocol indicated in the Procedure Manual. One examiner was a doctoral level licensed psychologist with 25 years of experience. The other examiner was completing her doctoral internship in school psychology. Each rater was provided with copies of the Procedure Manual, the Treatment Integrity Rating Instructions (See Appendix J)

and Treatment Integrity Rating Worksheets (See Appendix K). The investigator instructed the raters on rating procedures. The investigator calculated a kappa statistic to determine interjudge agreement as a measure of treatment integrity based on the raters' responses. High agreement between the raters was indicated by a kappa value of .911 ($p < .01$).

Treatments

Participants received an individual administration of the ten standard, non-supplementary WISC-III subtests under one of three treatment conditions (i.e., standard administration, verbal praise, or material reward). Subtest raw scores were determined by the examiners according to the scoring criteria found in the WISC-III Manual (Wechsler, 1991). Subtest and IQ scores were calculated by the investigator using the WISC-III Manual norm tables The experimental test procedures were as follows:

Standard Administration Condition

The participants received the standardized WISC-III administration as specified in the WISC-III procedure manual (Wechsler, 1991). Their efforts were not verbally or materially rewarded. They were given the subtests in a neutral, non-evaluative tone as described in the manual. During the administration of the 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. For every second response after the completion of the first three items, the examiner commented, "Give this one a try," "Let's try this," or "Let's try the next one." On the Coding subtest, the examiner provided comments when a participant completed a row of items. Between subtests, the examiner remarked, "Let's try something different." Examiner comments were stated regardless of the correctness of the participant's response or solution. In effect, the neutral comments controlled for verbal praise and material reinforcement that was administered according to the same fixed ratio, non-contingent schedule in their respective experimental conditions.

Verbal Praise Condition

Participants received the standardized WISC-III administration as specified in the WISC-III procedure manual (Wechsler, 1991) with one exception. Their efforts were verbally rewarded. During the administration of the subtests, the examiner stated, "Very good," "Fine," and "You are doing a very good job," after each of the first three responses in a subtest. For every second response after the completion of the first three items, the examiner commented, "Good job, keep it up," "Very good," or "Fine." The examiner provided comments for the Coding subtest when a participant completed a row of items. Between subtests, the examiner remarked, "That was very good. Let's try some more." Examiner

comments were stated regardless of the correctness of the participant's response or solution.

Material Reward Condition

Participants received the standardized WISC-III administration as specified in the WISC-III examiner's manual (Wechsler, 1991) with one exception. Their effort was requested and rewarded with tokens (i.e., pennies). The examiner dispensed the tokens following the same fixed ratio, noncontingent schedule used during the standard and verbal praise conditions (after each response on the first three items in a subtest; for every other response after the completion of the first three items; when a participant completed a Coding row; and between subtests). Following the completion of the test, a participant exchanged all the tokens for a pre-identified reward. The number of tokens earned by a participant varied, depending on the participant's starting point and discontinuation point (a minimum of 40 tokens and a maximum of 137 tokens).

Before the WISC-III administration, the examiner read the following statement to the participants in the material reward condition:

> You are going to have the opportunity to earn a reward for your efforts during testing. Your reward will be the preferred item that you identified for us on the Reward Menu. To do this, you must try your best on all items. If you try

hard, you should earn enough tokens to receive the reward that you want. You will need to earn at least 40 tokens to get your reward. You can trade your tokens in for your reward after the testing is completed.

The examiner dispensed the tokens regardless of the correctness of the response. The tokens were placed in a clear container by the examiner who made a neutral comment as the token was given (e.g., "Here", "Here you go", or "There").

If a participant made an effort to manipulate or count his or her tokens, the participant was reminded that the tokens would be tallied and exchanged for the reward after the testing was completed.

Reward Presentation

The participants in the standard administration and the verbal praise groups completed the WISC-III before the participants in the material reward group. This was done to avoid the distress that participants in the non-reward groups may experience as a result of not receiving rewards during testing. After the completion of the study, the participants in the verbal praise and the standard administration groups received rewards comparable to the participants in the material reward group to avoid disrupting the milieu of the residential/day treatment programs.

PostHoc Follow-up Questionnaire

After completing the WISC-III, all participants completed the Post Hoc Follow-up Questionnaire. Items from this questionnaire were read and explained to participants by the examiners if necessary.

Debriefing

The investigator met with each participant after all procedures were completed to inform participants about results, to answer questions, and to debrief participants. The investigator was prepared to arrange counseling for any participant who required this type of intervention; however, no participants identified the need for counseling related to their involvement in the study.

Design

The data was collected within the context of a posttest-only control group design (Campbell &Stanley, 1963) that controlled for threats against internal validity (history, maturation, testing, selection, instrumentation, statistical regression, experimental mortality, and selection-maturation interaction). Figure 2 presents a schematic representation of the data collection design.

Figure 2

Data Collection Design

	Treatment	Test
Random assignment	Material reward	WISC-III
Random assignment	Verbal praise	WISC-III
Random assignment	Standard condition	WISC-III

CHAPTER 3

RESULTS

This chapter presents the means and standard deviations for the WISC-III Verbal, Performance, and Full Scale IQ scores as well as the 10 standard WISC-III subtest scale scores for the treatment groups. The chapter also tests the hypotheses for this study. Results are presented in tables at the end of the chapter. Group means and standard deviations for the WISC-III Verbal, Performance, and Full Scale IQ scores as well as the 10 standard WISC-III subtest scale scores are presented in Table 8. Initially, the Hollingshead Index Composite scores were correlated with the WISC-III Full Scale IQs and a nonsignificant coefficient was observed ($\underline{r} = .176$, $\underline{p} = .17$). As such, it was decided that SES should not be used as a covariate in subsequent analyses.

The following hypotheses regarding the participants' WISC-III Full Scale and Verbal and Performance Scale IQ scores were tested:

H01: The mean Full Scale IQ score of the material reward group will be significantly greater than mean Full Scale IQ scores of the verbal praise and the control groups.

H02: The mean Verbal Scale IQ score of the material reward group will be significantly greater than mean Verbal Scale IQ scores of the verbal praise and the control groups.

H03: The mean Performance Scale IQ score of the material reward group will be significantly greater than mean Performance Scale IQ scores of the verbal praise and the control groups.

These hypotheses were tested using a multivariate analysis of variance (MANOVA) procedure. A nonsignificant value was observed using a Wilks' Lamba test, $\underline{F}$ (6,116) = .242, $\underline{p}$ = .91062. As such, hypotheses H01, H02, and H03 were not supported.

The following hypotheses regarding the participants' 10 standard WISC-III subtest scale scores were tested:

H04: The mean Picture Completion scaled score of the material reward group will be significantly greater than mean Picture Completion scaled scores of the verbal praise and the control groups.

H05: The mean Information scaled score of the material reward group will be significantly greater than mean Information scaled scores of the verbal praise and the control groups.

H06: The mean Coding scaled score of the material reward group will be significantly greater than mean Coding scaled scores of the verbal praise and the control groups.

H07: The mean Similarities scaled score of the material reward group will be significantly greater than mean Similarities scaled scores of the verbal praise and the control groups.

H08: The mean Picture Arrangement scaled score of the material reward group will be significantly greater than mean Picture Arrangement scaled scores of the verbal praise and the control groups.

H09: The mean Arithmetic scaled score of the material reward group will be significantly greater than the mean Arithmetic scaled scores of the verbal praise and the control groups.

H10: The mean Block Design scaled score of the material reward group will be significantly greater than mean Block Design scaled scores of the verbal praise and the control groups.

H11: The mean Vocabulary scaled score of the material reward group will be significantly greater than mean Vocabulary scaled scores of the verbal praise and the control groups.

H12: The mean Object Assembly scaled score of the material reward group will be significantly greater than mean Object Assembly scaled scores of the verbal praise and the control groups.

H13: The mean Comprehension scaled score of the material reward group will be significantly greater than mean Comprehension scaled scores of the verbal praise and the control groups

These hypotheses were tested using a MANOVA procedure. A nonsignificant value was observed using a Wilks' Lamba test, $\underline{F}$ (20,102) = 1.425, $\underline{p}$ = .128. As such, hypotheses H04, H05, H06, H07, H08, H09, H10, H11, H12, and H13 were not supported.

The following hypotheses regarding the participants' WISC-III Full Scale and Verbal and Performance Scale IQ scores were tested:

H14: The mean Full Scale IQ score of the verbal praise group will not be significantly different than the mean Full Scale IQ score of the control group.

H15: The mean Verbal Scale IQ score of the verbal praise group will not be significantly different than the mean Verbal Scale IQ score of the control group.

H16: The mean Performance Scale IQ score of the verbal praise group will not be significantly different than the mean Performance Scale IQ score of the control group.

These hypotheses were also tested using a MANOVA procedure. A nonsignificant difference was observed using a Wilks' Lamba test, $\underline{F}$ (6,116) = .242, $\underline{p}$ = .962. As such, hypotheses H14, H15, and H16 were supported.

The following hypotheses regarding the participants' 10 standard WISC-III subtest scale scores were tested:

H17: The mean Picture Completion scaled score of the verbal praise group will not be significantly different than the mean Picture Completion scaled score of the control group.

H18: The mean Information scaled score of the verbal praise group will not be significantly different than the mean Information scaled score of the control group.

H19: The mean Coding scaled score of the verbal praise group will not be significantly different than the mean Coding scaled score of the control group.

H20: The mean Similarities scaled score of the verbal praise group will not be significantly different than the mean Similarities scaled score of the control group.

H21: The mean Picture Arrangement scaled score of the verbal praise group will not be significantly different than the mean Picture Arrangement scaled score of the control group.

H22: The mean Arithmetic scaled score of the verbal praise group will not be significantly different than the mean Arithmetic scaled score of the control group.

H23: The mean Block Design scaled score of the verbal praise group will not be significantly different than the mean Block Design scaled score of the control group.

H24: The mean Vocabulary scaled score of the verbal praise group will not be significantly different than the mean Vocabulary scaled score of the control group.

H25: The mean Object Assembly scaled score of the verbal praise group will not be significantly different than the mean Object Assembly scaled score of the control group.

H26: The mean Comprehension scaled score of the verbal praise group will not be significantly different than the mean Comprehension scaled score of the control group.

These hypotheses were also tested using a MANOVA procedure. A nonsignificant difference was observed using a Wilks' Lamba test, F (20,102) = 1.425, p = .128). Given the nonsignificant F value, hypotheses H17, H18, H19, H20, H21, H22, H23, H24, H2512, and H26 were supported.

In addition, the following hypotheses regarding the participants' Post Hoc Follow-up Questionnaire total scores were tested:

H27: The Post Hoc Follow-up Questionnaire mean total score of the material reward group will be significantly greater than mean Post Hoc Follow-up Questionnaire total scores of the verbal praise and the control groups.

H28: The Post Hoc Follow-up Questionnaire total score of the verbal praise group will not be significantly different than the mean Post Hoc Follow-up Questionnaire total score of the control group.

Initially, means and standard deviations were calculated. These estimates are presented in Table 9. The hypotheses were tested using a one-way analysis of variance (ANOVA) and Bonferroni post hoc tests. The ANOVA denoted a significant difference between scores (F [2,60] = 13.695, p = .000). Bonferroni results for three pairwise combinations of conditions were completed. The mean score of the material reward group was significantly greater than the mean score

of the standard group (t [40] = 4.54, $p < .01$). The mean score of the verbal praise group was significantly greater than the mean score of the standard group (t [40] = 4.15, $p < .01$). The mean score of the material reward group was not significantly greater than the mean scores of the verbal praise group (t [40] = .74, $p = .46$). Given the observed results, none of the hypotheses for the Post Hoc Follow-up Questionnaire scores were supported.

Table 8

Group Means and Standard Deviations for the WISC-III Verbal, Performance, and Full Scale IQ Scores and the 10 Standard WISC-III Subtest Scale Scores

	Standard[a]		Verbal Praise[a]		Material Reward[a]	
Variable	Mean	SD	Mean	SD	Mean	SD
Verbal IQ	79.81	12.93	78.67	13.09	82.81	10.54
Performance IQ	84.43	15.50	83.10	14.26	87.00	14.64
Full Scale IQ	80.52	12.96	79.10	12.81	83.38	10.97
Information	6.90	3.02	6.24	2.59	6.71	2.49
Similarities	7.29	2.95	6.57	3.46	7.48	2.06
Arithmetic	5.52	2.58	6.52	2.36	6.48	2.36
Vocabulary	6.95	3.01	5.90	2.84	6.52	2.27
Comprehension	5.24	2.43	5.71	3.15	7.52	2.02
Picture Completion	8.43	3.04	8.62	3.06	9.05	1.96
Coding	6.71	3.10	5.52	2.80	6.95	3.06
Picture Arrangement	7.81	3.36	7.00	3.52	8.81	2.44
Block Design	6.90	4.12	6.76	2.57	7.10	4.10
Object Assembly	7.57	3.06	8.24	4.07	7.38	4.09

[a] $n = 21$ for each condition

Table 9

Mean Scores and Standard Deviations for the Post Hoc Follow-up Questionnaire

Condition	Mean	SD
Standard	54.29	8.38
Verbal Praise	63.81	6.35
Material Reward	65.38	7.43
Overall	61.16	8.83

CHAPTER 4

SUMMARY AND CONCLUSIONS

Summary

The present study compared the WISC-III performance of three groups of adolescents with conduct disorder under different incentive conditions (neutral comments, verbal praise, material reward). Youths with comorbid Attention-Deficit/Hyperactivity Disorder, Mood Disorders, Psychotic Disorders ,Substance Dependence, Dementia, Delirium, Mental Retardation, or Pervasive Developmental Disorder were excluded from the study. Similarly, participants who had a history of head trauma, seizure disorders, or epilepsy, or who were taking psychotropic medication were excluded from the study. A total of 427 adolescents, age range 12 – 16 years, were sampled from programs at two residential treatment centers. Sixty-three participants were selected for inclusion in the study from the sample of eligible youths that was randomly assigned to treatment conditions. The groups were comparable in respect to race/ethnicity, gender, age, and Hollingshead SES.

Independent MANOVAs determined that there were no significant differences between groups on the WISC-III Verbal, Performance, or Full Scale IQs. Likewise, nonsignificant group differences were evident across the WISC-III

subtests. A one-way ANOVA and Bonferroni tests determined that the material reward and verbal praise groups had significantly more positive thoughts about the testing experience relative to the control group. It was also determined that the self-reported affective reactions to the testing experience of the material reward and verbal praise groups were not significantly different.

As noted in Chapter 2, no examiner or examiner by gender effects were observed. Also as noted in Chapter 2, nonsignificant differences were observed for race/ethnicity, gender, and placement in residential treatment program.

Discussion and Implications

The nonsignificant differences observed between the material reward, verbal praise, and standard groups on the WISC-III performance do not support the hypothesis that adolescents with conduct disorder perform better on individually administered intelligence tests when given material rewards in comparison to groups who receive verbal praise or neutral comments. The nonsignificant results are inconsistent with the literature regarding the reinforcement preference of youth with conduct disorder (Quay, 1988a, 1988b, 1988c). It is recalled in this context that Saigh (1982) reported that conduct disorder youth who received material rewards evidenced significantly higher scores on a short form of the WISC-R relative to youth who received verbal praise or neutral comments. The

results are also inconsistent with Jackson et al.'s (1979) research that determined that material incentives improved the cognitive test performance of youth with severe behavioral problems. Likewise, these findings are inconsistent with the findings of very similar studies involving mildly retarded youth (Saigh & Payne, 1976; Terrell et al., 1981; Young et al., 1982). These studies consistently demonstrated that material rewards were associated with significantly higher WISC-R scores relative to verbal praise or neutral comments.

The results of this study should also be viewed with the realization that they may only apply to conduct disorder youth without comorbidity. The earlier incentive studies that reported improved test performance associated with material incentives (Saigh, 1982; Jackson et al., 1979) did not control for comorbiditiy. It is notable that children with conduct disorder who were comorbid for Attention-Deficit/Hyperactivity Disorder were responsive to material rewards on a response perseveration task (Daugherty & Quay, 1991). As such, the observed results may apply to youth with relatively "pure" conduct disorder.

The nonsignificant material incentive effects may have been due to the perceived value of the reinforcers that were used and the amount of effort that was needed to earn them. Examined in this context, the investigator offered $10 worth of tokens that were exchangeable for Chinese food, pizza, or a movie pass to

reinforce effort during the administration of the entire WISC-III. It should be noted that the WISC-III generally takes approximately one hour to administer and that participants may have viewed the reward as an insufficient motivator. The ineffectiveness of the material reward treatment may also have been due to relative satiation inasmuch as treats like pizza and movie passes were frequently available to the residential treatment participants on a noncontingent basis. It is important to note in this regard that Martin and Pear (1992) cautioned that " most reinforcers will not be effective unless the individuals have been deprived of them for some period of time prior to their use." (p. 36).

It is important to note that there is a general accord in the literature of individual intelligence testing regarding the need to facilitate optimal performance (Anastasi, 1997; National Association of School Psychologists, 1992; Sattler, 1992) during evaluations. In view of this, and as the self-ratings of the material reward and verbal praise groups were significantly greater than the control group's, these incentives may be an effective procedure for eliciting rapport and maintaining effort during the clinical evaluation of youth with conduct disorder.

Finally, the systematic use of verbal and material incentives is not in keeping with the WISC-III standardization procedures. Anastasi (1997) cautioned that influences that are specific to the test situation constitute error variance, affect test

validity, may limit or impair the generalizability of test results. In a similar vein, the individual attention provided by the examiners may have been perceived as rewarding and this may have influenced the test performance of the participants in all groups.

Limitations

The observed results must be tempered with the realization that selected participants were screened to control for the effects of a number of comorbid disorders. While this approach presents a good deal of theoretical merit, it should be understood that conduct disorder generally occurs with other disorders (e.g., Attention-Deficit/Hyperactivity Disorder; mood disorders; substance abuse). As such, the external validity of the observed results may be limited to comparison groups with similar demographic and clinical characteristics.

It is also possible that a different outcome may have been observed if different cognitive measures had been employed. The present findings may not apply to other measures of cognitive functioning that represent cognitive constructs that are different from the constructs assessed on the WISC-III. Material rewards may prove to be effective in improving the test performance of adolescents with conduct disorder on other cognitive measures, for example, the Wisconsin Card Sorting Test (Grant &Berg,1993) (WCST); the Stroop Color and Word Test

(Golden, 1978); and Raven's Progressive Matrices (Raven, 1965). Adolescents with conduct disorder may be more willing and able to sustain performance on these tasks as they involve significantly less testing time than the WISC-III. The participants may consider the amount of time and effort that is needed to complete these shorter tests as being commensurate with the rewards that were offered. Finally, the novelty of the tasks on the WCST, the Stroop Test, and Raven's Matrices may serve as a possible motivator in as much as all of the participants had received previous WISC-III administrations.

Recommendations

Given the observed results in the present study, a number of recommendations for future research are indicated. These recommendations are as follows:

1. In view of the nonsignificant differences between treatment conditions, it is recommended that future researchers conduct a similar study with a larger group of participants to provide greater power.
2. It is also recommended that potentially more valuable material rewards be used as incentives to overcome the possible motivational problems presented by youths with conduct disorder.
3. In view of the reported Conduct Disorder - Attention-Deficit/Hyperactivity Disorder comorbidity, it is recommended that a similar study be conducted with a this diagnostic group.
4. Alternatively, it is recommended that similar methodology be used with different dependent variables.

Notes

[1]Printed by permission from Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition,(pp. 90-91) by American Psychiatric Association, 1994, Washington, DC: American Psychiatric Association. Copyright 1994 by American Psychiatric Association.

Appendix Listing

Appendix Listing

Appendix A

Research Cover Sheet

RESEARCH COVER SHEET

NAME_____ PARTICIPANT NO. _____
CONSENT_____ ASSENT_____
DATE OF BIRTH_____ PROGRAM_____
DATE OF INTERVIEW_____ GRADE___RES.STATUS_____
AGE_____GENDER_____ ETHNICITY_____

FATHER AT HOME: YES/NO MOTHER AT HOME: YES/NO
FATHER OCCUP _____ MOTHER OCCUP_____
FATHER EDUC _____ MOTHER EDUC _____
FAMILY MAKEUP _____
GEOGRAPHIC LOCATION OF RESIDENCE: URB____SUB____RUR____

HOLLINGSHEAD SES SCORE _____

DICA ADHD: YES/NO PSYCHOTROPIC MEDS: YES/NO
DICA CD : YES/NO MR/PDD: YES/NO
DICA MOOD DO : YES/NO SIGNIF HEAD TRAUMA: YES/NO
DICA SUB DEP : YES/NO EPILEPSY/SEIZURE DO: YES/NO
DICA PSYCHOT : YES/NO DELIRIUM/DEMENTIA: YES/NO
RECENT WISC-III: YES/NO

IDENTIFIED REWARD_____(MED/CULT CONSTR: YES/NO)

WISC-III SCORES

VERBAL SUBTESTS		PERFORMANCE SUBTESTS	
Information	__	Picture Completion	__
Similarities	__	Coding	__
Arithmetic	__	Picture Arrangement	__
Vocabulary	__	Block Design	__
Comprehension	__	Object Assembly	__

VERBAL IQ ___ PERFORMANCE IQ ___ FULL SCALE IQ ___

RESEARCH COVER SHEET

POST HOC FOLLOWUP QUESTIONNAIRE TOTAL SCORE ___

DEBRIEFING SESSION: _____

KAPPA SELECTION: YES/NO

Appendix B

Reward Menu

REWARD MENU

NAME:_____
DATE:__/__/__

Listed below are some items which students find rewarding and have a chance to earn in their classrooms and residences.

List the five items you find most rewarding on the lines below.

If your favorite rewards are not on the list, you can think of similar rewards that you prefer and write them down on the lines below.

MOVIE PASS	MCDONALDS COUPONS
CHINESE FOOD	TELEPHONE CARD
PIZZA	MAGAZINES
ICE CREAM	COMIC BOOKS
BATTERIES	"NERF" FOOTBALL
CASSETTE TAPES	MAKE-UP
HEADBANDS	DISPOSABLE CAMERA
T-SHIRT	PUZZLES
CAP	STUFFED ANIMALS

REWARD MENU

FIRST CHOICE: ----------

SECOND CHOICE: ----------

THIRD CHOICE: ----------

FOURTH CHOICE: ----------

FIFTH CHOICE:----------

Appendix C

Post Hoc Follow-up Questionnaire

POST HOC FOLLOW-UP QUESTIONNAIRE

Please read the following statements and rate each statement based on your experience during the testing.

Indicate by circling the appropriate letter if you strongly agree, agree, disagree, strongly disagree with the statement.

A Strongly Agree **B** Agree **C** Disagree **D** Strongly Disagree

1. While taking the test, I thought I was doing well.	A	B	C	D
2. This test was interesting.	A	B	C	D
3. I did my best on this test.	A	B	C	D
4. I gave up easily on the test.	A	B	C	D
5. I paid attention to the test questions.	A	B	C	D
6. I tried hard to answer all test questions.	A	B	C	D
7. Taking this test made me worry a lot.	A	B	C	D
8. I was able to concentrate on the test.	A	B	C	D
9. I tried hard to do well on the test.	A	B	C	D
10. I felt relaxed during the test.	A	B	C	D
11. I did not mind answering questions out loud during the test.	A	B	C	D
12. I told myself to try hard during the test.	A	B	C	D

POST HOC FOLLOW-UP QUESTIONNAIRE

Please read the following statements and rate each statement based on your experience during the testing.

Indicate by circling the appropriate letter if you strongly agree, agree, disagree, strongly disagree with the statement.

A Strongly Agree **B** Agree **C** Disagree **D** Strongly Disagree

13. Taking this test made me less willing to take tests in the future.	A	B	C	D
14. As the test went on, I found myself willing to try harder to do well on the test.	A	B	C	D
15. Examiner comments during the testing distracted me.	A	B	C	D
16. Taking the test made me feel like I don't know much.	A	B	C	D
17. My level of effort was good on the test from beginning to end.	A	B	C	D
18. Examiner comments during the test made me feel uncomfortable.	A	B	C	D
19. Taking this test will help me be more willing to take tests in the future.	A	B	C	D
20. Most people would be willing to take a test like this test.	A	B	C	D

Appendix D

Hollingshead Four Factor Index of Social Status Worksheet

HOLLINGSHEAD FOUR FACTOR INDEX OF SOCIAL STATUS

WORKSHEET

Subject Number: _____ Date: ___/___/___
Mother at home? YES/NO Father at home? YES/NO
Mother's occupation_____ Father's occupation_____
Occupation Scaled Score____ Occupation Scaled Score____

CIRCLE ONE:

Mother's education:	Score	Father's education:	Score
< 7th grade	1	< 7th grade	1
9th grade	2	9th grade	2
10th or 11th grade	3	10th or 11th grade	3
HS graduate	4	HS graduate	4
> 1 yr. of college	5	> 1 yr. of college	5
college/univ. grad.	6	college/univ. grad.	6
graduate degree	7	graduate degree	7

Head of household married? YES/NO
Separated or divorced? YES/NO
If separated-divorced, does head receive support payments? YES/NO

Father's scoring:

Factor	Scaled Score	Factor Weight	Score x Weight
Occupation	____	__5__	_____
Education	____	__3__	_____

Mother's scoring:

Factor	Scaled Score	Factor Weight	Score x Weight
Occupation	____	__5__	_____
Education	____	__3__	_____

If appropriate:
Father's Score ____
Mother's Score ____
Total Scores ____ divided by 2 = ____
Family Total Score

Appendix E

Procedure Manual: The Effects of Different Incentive Conditions on the WISC-III Performance of Conduct Disorder Adolescents

PROCEDURE MANUAL

The Effects of Different Incentive Conditions on the WISC-III Performance of Conduct Disorder Adolescents

Residents/students between the ages 12 through 16 years from a residential/day treatment program will be invited to participate in the study and will be randomly assigned to three treatment groups. The eligibility of 63 adolescents with conduct disorder will be determined, and the eligible participants will be serially selected for each group (n = 21 per group).

APPROVAL, CONSENT, ASSENT

Agency permission will be obtained through the Agency's human subjects review board for the research project.

Agency staff who will be contacted during the research project will be informed about the project in writing.

A list of potential participants, ages 12 through 16, will be compiled from Agency data sources.

Based on a review of records:

- participants who are taking psychotropic medication will be excluded from the study.
- participants who have been administered the WISC-III within the past 12 months will be excluded from the study.
- rewards which are medically or culturally contra-indicated will be identified for each participant.

Each eligible participant will be randomly assigned to one of three groups.

The parents or legal guardians of each participant will receive a cover letter and a consent form (See Appendices A and B for examples from one agency) for the participant. Spanish translations of the cover letter and consent form will be available if necessary. Translations into other languages will be completed if necessary. Follow-up calls and mail contacts will be made if necessary.

PROCEDURE MANUAL

The Effects of Different Incentive Conditions on the WISC-III Performance of Conduct Disorder Adolescents

Participant assent (See Appendix C for an example from one agency) will be obtained in writing from participants who have parent/guardian permission to take part in the study. Participants will be informed that they are free to withdraw from the study at any time without reprimand.

DATA COLLECTION

The investigator will administer the Conduct Disorder Module from the DICA to each participant in order to identify participants who meet DSM-IV criteria for Conduct Disorder. Information from the participant's records will be used to corroborate the diagnosis.

The,DICA Attention-Deficit/Hyperactivity Disorder, Mood Disorders, Psychotic Disorders, and Substance Dependence Modules, as well as corroborating information from observations and record reviews will be used to exclude participants with comorbid Attention-Deficit/ Hyperactivity Disorder, Mood Disorders, Psychotic Disorders, Substance Dependence, Dementia, Delirium, Mental Retardation, or Pervasive Developmental Disorder. Participants who have a history of head trauma, seizure disorders, or epilepsy, will be excluded from the study.

Information regarding effective and appropriate reinforcers for each participant will be obtained by asking each participant to indicate his or her preferences for reinforcers on the Reward Menu (See Appendix D). Each participant's preference for an item or items worth $10.00 will be elicited.

The investigator will collect information from each participant's records and interview responses regarding parental education and occupation. This information will be used to establish the Hollingshead (See Appendix E) ratings of the participants.

PROCEDURE MANUAL

The Effects of Different Incentive Conditions on the WISC-III Performance of Conduct Disorder Adolescents

Participants will be assigned to one of six examiners who hold current New York State School Psychologist certification for testing based on examiner availability at the time when a participant becomes eligible and available for the study. If more than one examiner is available to work with a participant, the participant will be randomly assigned to an examiner. Examiners will be informed that they must abide by all federal, state, and local laws governing ethical research with vulnerable populations and that they are mandated reporters of child abuse and neglect.

Participants will receive an individual administration of the WISC-III under one of three treatment conditions (i.e., standard administration, verbal praise, or material reward). The WISC-III will be administered to all participants according to specific directions stated on Test Procedure sheets prepared for each condition (See Appendices F, G, and H).

All WISC-III testing will be recorded on audiotape (unless a participant strongly objects and refuses to be recorded).

DATA COLLECTION

Administering rewards to a select group in a residential setting is likely to establish expectations of receiving rewards during testing for participants in all groups. To avoid the confounding expectations and the associated distress caused by not receiving rewards during testing, the participants in the Standard Administration and the Verbal Praise Groups will complete the WISC-III before the participants in the Material Reward Group. The participants in the Standard Administration and the Verbal Praise Groups will receive their item of preference from the Reward Menu after testing with both groups is completed and before testing of the Material Reward Group occurs.

Subtest raw scores will be determined by the examiner according to the WISC-III Manual.

PROCEDURE MANUAL

The Effects of Different Incentive Conditions on the WISC-III Performance of Conduct Disorder Adolescents

Subtest and IQ scores will be calculated by the investigator using the WISC-III Manual norm tables.

After completing the WISC-III, all participants will complete the Post Hoc Follow-up Questionnaire (See Appendix I). Items from this questionnaire will be read and explained to participants by the examiner if necessary.

All test scores and interview results will be confidential and stored in a locked file by the investigator.

After each test administration, the investigator will enter data from the Research Cover Sheet (See Appendix J) to a computer disc that will not include the participants' names for data analysis procedures

POST DATA COLLECTION FOLLOW-UP

The investigator will meet with each participant after all procedures have been completed to inform participants about results, to answer questions, and to debrief participants.

The investigator will arrange for counseling for any participant who requires that type of intervention.

DATA ANALYSIS

The WISC-II data obtained from the three groups will be analyzed using a multiple analysis of covariance procedure (MANOVA).

Two,examiners will independently rate the tape-recorded WISC-III administration for five randomly selected participants from each group for compliance with the experimental protocol indicated in the Procedure Manual for the study (See Appendix K for copy of the Treatment Integrity Rating Instructions

PROCEDURE MANUAL

The Effects of Different Incentive Conditions on the WISC-III Performance of Conduct Disorder Adolescents

and Appendix L for copy of the Treatment Integrity Rating Worksheet). A kappa statistic will be obtained to determine treatment integrity.

Appendix F

Test Procedures

TEST PROCEDURES

STANDARD ADMINISTRATION GROUP

The participants in the Standard Administration Group will receive the **standardized WISC-III administration and scoring as specified in the WISC-III procedure manual** (Wechsler, 1991). They will be given **subtests one through ten** in a neutral, non-evaluative manner as described in the manual. Supplementary subtests will not be administered.

- Standardized procedures include using a natural, nonthreatening, conversational tone, encouraging interest in tasks, and reinforcing the child's effort.
- Testing should proceed at a steady pace. Brief conversations between subtests may help to maintain cooperation and interest and reduce test apprehension.
- Short breaks can be provided if necessary and should occur at the completion of a subtest.
- Every effort to administer the entire test in a single session should be made. Fatigue, inadequate motivation, or other reasons may necessitate discontinuation and rescheduling of a second session within the time period of one week.
- Efforts should be made to minimize any potential distraction or interference. The physical setting should be quiet, adequately lit, and well ventilated.
- Seating arrangements and organization of materials should allow easy access to test materials, promote the child's comfort and ease of manipulating the materials, keep the materials out of the child's view until they are needed, and allow an unobstructed view of the child's responses and behaviors.
- Feedback on whether a particular response is right or wrong should not be given.

The examiner will **audiotape the test session** unless the participant strongly objects. (If the participant objects, note it on the test protocol.)

Introduce testing by informing the child that:

- He or she will be asked to engage in tasks that most young people enjoy.
- Some of the tasks will be easy while others may be more difficult.

TEST PROCEDURES

STANDARD ADMINISTRATION GROUP

- He or she may not answer all of the questions correctly.
- He or she is expected to try their best on all items.

During the administration of the subtests:

After each response on the first three items in a subtest the examiner will state:

"Now try this." (first item)
"How about this." (second item)
"Here is the next one." (third item)

For every other response (ie., fifth, seventh, etc.) after the completion of the first three items, the examiner will comment on an alternating basis:

"Give this one a try." (e.g., fifth item)
"Let's try this." or (e.g., seventh item)
"Let's try the next one." (e.g., ninth item)

On the Coding subtest, the examiner will provide the comments listed below on an alternating basis when a participant completes a Coding row:

"Give this one a try" (end of first row)
"Let's try this" or (end of second row)
"Let's try the next one." (end of third row)

Between subtests, the examiner will remark:

"Let's try something different."

The examiner will note each statement on the test protocol by **marking** the item, row, or subtest **with an asterisk**.

After the test is completed, the examiner will present the **Post Hoc Follow-up Questionnaire** for completion by the participant. Items from this questionnaire will be read and explained to the participant by the examiner if necessary.

TEST PROCEDURES

STANDARD ADMINISTRATION GROUP

After all procedures are completed, the examiner will inform the participant that the investigator will meet with him or her to discuss the test results and procedures.

TEST PROCEDURES

VERBAL PRAISE GROUP

The participants in the Verbal Praise Group will receive the **standardized WISC-III administration and scoring as specified in the WISC-III procedure manual** (Wechsler, 1991) with one exception. Their effort in completing an item will be verbally rewarded as described below. They will be given **subtests one through ten**. Supplementary subtests will not be administered.

- Standardized procedures include using a natural, nonthreatening, conversational tone, encouraging interest in tasks, and reinforcing the child's effort.
- Testing should proceed at a steady pace. Brief conversations between subtests may help to maintain cooperation and interest and reduce test apprehension.
- Short breaks can be provided if necessary and should occur at the completion of a subtest.
- Every effort to administer the entire test in a single session should be made. Fatigue, inadequate motivation, or other reasons may necessitate discontinuation and rescheduling of a second session within the time period of one week.
- Efforts should be made to minimize any potential distraction or interference. The physical setting should be quiet, adequately lit, and well ventilated.
- Seating arrangements and organization of materials should allow easy access to test materials, promote the child's comfort and ease of manipulating the materials, keep the materials out of the child's view until they are needed, and allow an unobstructed view of the child's responses and behaviors.
- Feedback on whether a particular response is right or wrong should not be given.

The examiner will **audiotape the test session** unless the participant strongly objects. (If the participant objects, note it on the test protocol.)

Introduce testing by informing the child that:

- He or she will be asked to engage in tasks that most young people enjoy.
- Some of the tasks will be easy while others may be more difficult.

TEST PROCEDURES

VERBAL PRAISE GROUP

- He or she may not answer all of the questions correctly.
- He or she is expected to try their best on all items.

During the administration of the subtests:

After each response on the first three items in a subtest, the examiner will state:

"Very good." (first item)
"Fine." (second item)
"You are doing a very good job." (third item)

For every other response (ie., fifth, seventh, etc.) after the completion of the first three items, the examiner will comment on an alternating basis:

"Good job, keep it up." (e.g., fifth item)
"Very good." or (e.g., seventh item)
"Fine." (e.g., ninth item)

On the Coding subtest, the examiner will provide the comments listed below on an alternating basis when a participant completes a Coding row:

"Good job, keep it up." (end of first row)
"Very good." or (end of second row)
"Fine." (end of third row)

Between subtests, the examiner will remark:

"That was very good. Let's try some more."

The examiner will note each statement on the test protocol by **marking** the item, row, or subtest **with an asterisk**.

After the test is completed, the examiner will present the **Post Hoc Follow-up Questionnaire** for completion by the participant. Items from this questionnaire will be read and explained to the participant by the examiner if necessary.

TEST PROCEDURES

VERBAL PRAISE GROUP

After all procedures are completed, the examiner will inform the participant that the investigator will meet with him or her to discuss the test results and procedures.

TEST PROCEDURES

MATERIAL REWARD GROUP

The participants in the Material Reward Group will receive the **standardized WISC-III administration and scoring as specified in the WISC-III procedure manual** (Wechsler, 1991) with one exception. Their effort in completing an item will be rewarded with tokens (i.e., pennies). They will be given **subtests one through ten**. Supplementary subtests will not be administered. The investigator will provide tokens and the pre-identified reinforcer.

- Standardized procedures include using a natural, nonthreatening, conversational tone, encouraging interest in tasks, and reinforcing the child's effort.
- Testing should proceed at a steady pace. Brief conversations between subtests may help to maintain cooperation and interest and reduce test apprehension.
- Short breaks can be provided if necessary and should occur at the completion of a subtest.
- Every effort to administer the entire test in a single session should be made. Fatigue, inadequate motivation, or other reasons may necessitate discontinuation and rescheduling of a second session within the time period of one week.
- Efforts should be made to minimize any potential distraction or interference. The physical setting should be quiet, adequately lit, and well ventilated.
- Seating arrangements and organization of materials should allow easy access to test materials, promote the child's comfort and ease of manipulating the materials, keep the materials out of the child's view until they are needed, and allow an unobstructed view of the child's responses and behaviors.
- Feedback on whether a particular response is right or wrong should not be given.

The examiner will **audiotape the test session** unless the participant strongly objects. (If the participant objects, note it on the test protocol.)

Introduce testing by informing the child that:

- He or she will be asked to engage in tasks that most young people enjoy.

TEST PROCEDURES

MATERIAL REWARD GROUP

- Some of the tasks will be easy while others may be more difficult.
- He or she may not answer all of the questions correctly.
- He or she is expected to try their best on all items.

Before the WISC-III administration, the examiner will read the following statement to each participant in the material reward condition:

> **You are going to have the opportunity to earn a reward for your efforts during testing. Your reward will be the preferred item that you identified for us on the Reward Menu. To do this, you must try your best on all items. If you try hard, you should earn enough tokens to receive the reward that you want. You will need to earn at least 40 tokens to get your reward. You can trade your tokens in for your reward after the testing is completed.**

Tokens will be placed in a clear container by the examiner who will make a neutral comment as the token is given (e.g., "Here", "Here you go", or There").

The examiner will dispense the a token and verbalize the statement as follows regardless of the correctness of the response:

- **After each response on the first three items in a subtest**
- **For every other response (i.e., fifth, seventh, etc.) after the completion of the first three items**
- **On the Coding subtest, the examiner will provide a token when a participant completes a Coding row**
- **Between subtests**

The examiner will note each dispensed token on the test protocol by **marking** the item, row, or subtest **with an asterisk.**

TEST PROCEDURES

MATERIAL REWARD GROUP

If a participant makes an effort to manipulate or count his or her tokens, the participant will be reminded **"You can count and trade your tokens in for your reward after the testing is completed."**

If the participant inquires about how many tokens are needed to receive the reward, the examiner should state, **"If you try hard, you should earn enough tokens to receive the reward that you want."**

Following the completion of the test, a participant will exchange all the tokens for the **pre-identified reinforcer** regardless of the number of tokens accumulated. (The number of tokens earned by a participant will vary depending on the participant's starting point and discontinuation point up.)

After the test is completed, the examiner will present the **Post Hoc Follow-up Questionnaire** for completion by the participant. Items from this questionnaire will be read and explained to the participant by the examiner if necessary.

After all procedures are completed, the examiner will inform the participant that the investigator will meet with him or her to discuss the test results and procedures.

Appendix G

Cover Letter to Parents

The Graduate School and University Center
The City University of New York
Ph.D. Program in Educational Psychology
365 Fifth Avenue
New York, New York 10016-4309
212-817-8285

(Date), 2000

Dear Parent(s) or Guardian(s):

I am a student at the City University of New York Graduate School and University Center. As part of my studies in the area of School Psychology, I am completing a project that will try to find ways to help students during testing. The (Agency name) has agreed to allow (Program name) to be a study site. The attached information outlines the purpose and procedures of the study.

If you are willing to have your child participate, please indicate your approval on the attached consent form, sign the form, and return it to me in the enclosed envelope. You are encouraged to discuss this decision with your child. If you agree to allow your child to participate, I will explain the study to him or her and obtain assent.

If you decide not to have your child participate, please indicate your disapproval on the attached consent form, sign the form, and return it to me in the enclosed envelope. Your child will not be penalized in any way.

If you have any questions, please feel free to call me at 914-375-8627.

Thank you for considering my request.

Sincerely,

Ed Fallon M.S.
CUNY Doctoral Student

Appendix H

Consent Form

The Graduate School and University Center
The City University of New York
Ph.D. Program in Educational Psychology
365 Fifth Avenue
New York, New York 10016-4309
212-817-8285

PARENT/ GUARDIAN CONSENT FORM

STUDY TITLE: The Effects of Different Incentive Conditions on the WISC-III Performance of Adolescents with Conduct Disorder

INVESTIGATOR: Ed Fallon, M.S., Doctoral Student

PURPOSE: To find ways to help students during testing.

PROCEDURES: The project will consist of two parts :

Day 1: An interview to gather information about behavior that may affect testing (30 to 60 minutes).

Day 2: An individually administered intelligence test routinely taken by students (60 to 90 minutes).
Testing will not result in missing school work. Mr. Fallon will visit students to answer questions after the study is completed.

RISKS: Participants will be exposed to minimal risks. No procedures will occur without the written consent of the parents or guardians and the child. Children must also voluntarily consent to participate in the study and sign an assent form after an explanation of the study. Any child who changes his or her mind and decides not to participate may withdraw from the project without penalty. All test scores and records will be kept confidential and stored in a locked file.

BENEFITS: Students who participate may feel better about taking tests in school. Students who have completed similar activities in the past appeared to enjoy the individual attention and interesting test materials. In addition, participation may help other young people who need to take tests in school. If it appears that your child is experiencing distress during the study, counseling will be arranged.

PARENT/ GUARDIAN CONSENT FORM

QUESTIONS: If you have any questions about the project, contact Mr. Fallon at 914-375-8627, Professor Philip A. Saigh at CUNY 212-817-8285, or Hilry Fisher, Sponsored Research, CUNY/GSUCenter at 212-817-7525.

I have read the description of the study and have had the option to discuss the procedures with my child.
_____ I give consent _____ I do not give for my child to participate in the study.

_____Date_____ _____ _____ Date _____

(Signature of Parent[s] or Guardian[s])

Student: _____ Investigator: Ed Fallon M.S. Date_____

Appendix I

Participant Assent Form

The Graduate School and University Center
The City University of New York
Ph.D. Program in Educational Psychology
365 Fifth Avenue
New York, New York 10016-4309
212-817-8285

PARTICIPANT ASSENT FORM

Study Title: The Effects of Different Incentive Conditions on the WISC-III Performance of Adolescents with Conduct Disorder

Investigator: Edmond Fallon, M.S., Doctoral Student

I am being asked to help Mr. Fallon in a project. The goal of this project is to find out more about ways to help students during testing. If I decide to participate, my part in the project will involve two activities:

Day 1 - I will be interviewed about different types of behavior that may affect testing. This will take about 30 to 60 minutes.

Day 2 - I will take an individually administered test routinely taken by students. This will take about 60 to 90 minutes.

I have been informed that:

- The activities will not result in missing school.
- After all of the students have been tested, the investigator will visit me to answer questions.
- If I decide to participate, I will be exposed to minimal risks.
- No interviewing or testing will take place without the consent of my parents or guardians and myself.
- All information will be kept confidential and records will be stored in a locked file.

PARTICIPANT ASSENT FORM

- Taking part in this project may make me feel better about taking tests.
- Other students who have completed similar activities have enjoyed the activities.
- This study will provide information that may help other students who take tests.
- Counseling will be arranged if I experience distress during this project.

QUESTIONS

If I have any questions about the project, I can contact Mr. Fallon at 914-375-8627 or Professor Philip A. Saigh at CUNY 212-817-8285. If I have any questions concerning my rights as a participant in this study, I can call Hilry Fisher, Sponsored Research, Graduate School and University Center/CUNY at 212-817-7525.

ASSENT

This project has been explained to me and I have been allowed to ask questions about it. I understand that I do not have to participate if I don't want to and no one will treat me badly. I can stop part way through if I want to and skip questions I don't want to answer. I have had the option to discuss the procedures with my parents or guardian. I have read this form, understand the project, and agree to participate.

--- I give my assent to participate in the study.

--- I do not give my assent to participate in the study.

---------- ----------

(Signature of Student) (Date)

---------- ----------

Ed Fallon M.S., Investigator) (Date)

Appendix J

Treatment Integrity Rating Instructions

The Effects of Different Incentive Conditions on the WISC-III Performance of *Conduct Disorder Adolescents*

Treatment Integrity Rating Instructions

PURPOSE:

Two,examiners will independently rate the tape-recorded WISC-III administration for 15 randomly selected participants (five from each experimental condition) for compliance with the experimental protocol indicated in the Procedure Manual for the study. The investigator will obtain a kappa statistic to determine treatment integrity.

MATERIALS:

- Procedure Manual including Test Procedures for the Standard Administration Group, the Verbal Praise Group, and the Material Reward Group

- Treatment Integrity Rating Worksheets for each selected participant. Items administered by the examiners in accordance with standard WISC-III procedures have been indicated for each participant. Items scheduled for treatment in accordance with the test procedures for this study have also been indicated for each participant.

- Cassette audiotapes of the WISC-III administration for each selected participant.

- WISC-III protocol for each selected participant (to assist the rater in listening to the tape).

PROCEDURES:

- Raters will indicate if a standard WISC-III introduction statement or the Material Reward Condition WISC-III introduction statement was made by the examiner (per guidelines or statement provided in Procedure Manual - Test Procedures sections).

Treatment Integrity Rating Instructions

- Raters will indicate if the required statements or rewards for subtest items occurred for items scheduled to receive statements or rewards (per instructions provided in Procedure Manual - Test Procedures sections). Raters can accept any one of the statements listed for the respective conditions. (Changes in the sequence of statements, if they occur, are not to be rated as deviations from treatment procedures within a treatment condition. Statements for each treatment are considered to be equivalent for their respective treatment.)

- Raters will indicate if the required statements or rewards occurred between subtests (per instructions provided in Procedure Manual - Test Procedures sections). Again, raters can accept any one of the statements listed for the respective conditions. (Changes in the sequence of statements, if they occur, are not to be rated as deviations from treatment procedures within a treatment condition. Statements for each treatment are considered to be equivalent for their respective treatment.)

- Raters will indicate if the WISC-III was administered in the standard manner as specified in the WISC-III procedure manual (Wechsler, 19991) except for changes associated with the Verbal Praise and Material Reward treatments (per instructions provided in Procedure Manual - Test Procedures sections). Violations of standard administration should be obvious and extreme (eg., statements indicating that a response is right or wrong).

- If the quality of the tape recording is poor, or if the examiner's or participant's responses are unintelligible or inaudible for some items, raters will indicate their inability to rate the item by marking the item "UR" next to the "YES/NO" response which could not be made. (If 80% of the items are unable to be rated, another participant will be selected at random to replace the participant whose response could not be rated adequately.)

Appendix K

Treatment Integrity Rating Worksheets

The Effects of Different Incentive Conditions on the WISC-III Performance of Conduct Disorder Adolescents

Treatment Integrity Rating Worksheet

Standard Administration Condition

Participant Selected at Random:

Rater: _____
Date: _____

WECHSLER INTELLIGENCE SCALE FOR CHILDREN – THIRD EDITION

STANDARD WISC-III INTRODUCTION STATEMENT	YES	NO

1. PICTURE COMPLETION SUBTEST

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO

18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO
25	YES / NO	YES / NO	YES / NO
26	YES / NO	YES / NO	YES / NO
27	YES / NO	YES / NO	YES / NO
28	YES / NO	YES / NO	YES / NO
29	YES / NO	YES / NO	YES / NO
30	YES / NO	YES / NO	YES / NO

2. INFORMATION SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO

18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO
25	YES / NO	YES / NO	YES / NO
26	YES / NO	YES / NO	YES / NO
27	YES / NO	YES / NO	YES / NO
28	YES / NO	YES / NO	YES / NO
29	YES / NO	YES / NO	YES / NO
30	YES / NO	YES / NO	YES / NO

3. CODING SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ROW	ROW ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO

4. SIMILARITIES SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO

6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO

5. PICTURE ARRANGEMENT SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1 trial 1	YES / NO	YES / NO	YES / NO
1 trial 2	YES / NO	YES / NO	YES / NO
2 trial 1	YES / NO	YES / NO	YES / NO
2 trial 2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO

6. ARITHMETIC SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO

7. BLOCK DESIGN SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1 trial 1	YES / NO	YES / NO	YES / NO

1 trial 2	YES / NO	YES / NO	YES / NO
2 trial 1	YES / NO	YES / NO	YES / NO
2 trial 2	YES / NO	YES / NO	YES / NO
3 trial 1	YES / NO	YES / NO	YES / NO
3 trial 2	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11*	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO

8. VOCABULARY SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO

17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO
25	YES / NO	YES / NO	YES / NO
26	YES / NO	YES / NO	YES / NO
27	YES / NO	YES / NO	YES / NO
28	YES / NO	YES / NO	YES / NO
29	YES / NO	YES / NO	YES / NO
30	YES / NO	YES / NO	YES / NO

9. OBJECT ASSEMBLY SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO

10. COMPREHENSION SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO

6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO

STANDARD WISC-III ADMINISTRATION WAS MAINTAINED	YES	NO

TOTAL STATEMENTS SCHEDULED PER TREATMENT PROCEDURE	
TOTAL STATEMENTS MADE PER RATER OBSERVATION	

The Effects of Different Incentive Conditions on the WISC-III Performance of Adolescents with Conduct Disorder

Treatment Integrity Rating Worksheet

Verbal Praise Condition

Participant Selected at Random:

Rater: _____
Date: _____

WECHSLER INTELLIGENCE SCALE FOR CHILDREN – THIRD EDITION

STANDARD WISC-III INTRODUCTION STATEMENT	YES	NO

1. PICTURE COMPLETION SUBTEST

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO

18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO
25	YES / NO	YES / NO	YES / NO
26	YES / NO	YES / NO	YES / NO
27	YES / NO	YES / NO	YES / NO
28	YES / NO	YES / NO	YES / NO
29	YES / NO	YES / NO	YES / NO
30	YES / NO	YES / NO	YES / NO

2. INFORMATION SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO

18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO
25	YES / NO	YES / NO	YES / NO
26	YES / NO	YES / NO	YES / NO
27	YES / NO	YES / NO	YES / NO
28	YES / NO	YES / NO	YES / NO
29	YES / NO	YES / NO	YES / NO
30	YES / NO	YES / NO	YES / NO

3. CODING SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ROW	ROW ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO

4. SIMILARITIES SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO

6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO

5. PICTURE ARRANGEMENT SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1 trial 1	YES / NO	YES / NO	YES / NO
1 trial 2	YES / NO	YES / NO	YES / NO
2 trial 1	YES / NO	YES / NO	YES / NO
2 trial 2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO

6. ARITHMETIC SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO

7. BLOCK DESIGN SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1 trial 1	YES / NO	YES / NO	YES / NO
1 trial 2	YES / NO	YES / NO	YES / NO
2 trial 1	YES / NO	YES / NO	YES / NO
2 trial 2	YES / NO	YES / NO	YES / NO
3 trial 1	YES / NO	YES / NO	YES / NO
3 trial 2	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO

8. VOCABULARY SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO

11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO
25	YES / NO	YES / NO	YES / NO
26	YES / NO	YES / NO	YES / NO
27	YES / NO	YES / NO	YES / NO
28	YES / NO	YES / NO	YES / NO
29	YES / NO	YES / NO	YES / NO
30	YES / NO	YES / NO	YES / NO

9. OBJECT ASSEMBLY SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO

10. COMPREHENSION SUBTEST

SCHEDULED STATEMENT MADE BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	STATEMENT SCHEDULED	STATEMENT MADE
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO

STANDARD WISC-III ADMINISTRATION WAS MAINTAINED (EXCEPT FOR VERBAL PRAISE TREATMENT)	YES	NO

STATEMENTS SCHEDULED PER TREATMENT PROCEDURE	
STATEMENTS MADE PER RATER OBSERVATION	

The Effects of Different Incentive Conditions on the WISC-III Performance of Adolescents with Conduct Disorder

Treatment Integrity Rating Worksheet

Material Reward Condition

Participant Selected at Random:

Rater: _____
Date: _____

WECHSLER INTELLIGENCE SCALE FOR CHILDREN – THIRD EDITION

MATERIAL REWARD CONDITION WISC-III INTRODUCTION STATEMENT	YES	NO

1. PICTURE COMPLETION SUBTEST

ITEM	ITEM ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO

17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO
25	YES / NO	YES / NO	YES / NO
26	YES / NO	YES / NO	YES / NO
27	YES / NO	YES / NO	YES / NO
28	YES / NO	YES / NO	YES / NO
29	YES / NO	YES / NO	YES / NO
30	YES / NO	YES / NO	YES / NO

2. INFORMATION SUBTEST

SCHEDULED REWARD GIVEN BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO

17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO
25	YES / NO	YES / NO	YES / NO
26	YES / NO	YES / NO	YES / NO
27	YES / NO	YES / NO	YES / NO
28	YES / NO	YES / NO	YES / NO
29	YES / NO	YES / NO	YES / NO
30	YES / NO	YES / NO	YES / NO

3. CODING SUBTEST

SCHEDULED REWARD GIVEN BETWEEN SUBTESTS	YES	NO

ROW	ROW ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO

4. SIMILARITIES SUBTEST

SCHEDULED REWARD GIVEN BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO

5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO

5. PICTURE ARRANGEMENT SUBTEST

SCHEDULED REWARD GIVEN BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1 trial 1	YES / NO	YES / NO	YES / NO
1 trial 2	YES / NO	YES / NO	YES / NO
2 trial 1	YES / NO	YES / NO	YES / NO
2 trial 2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO

14	YES / NO	YES / NO	YES / NO

6. ARITHMETIC SUBTEST

SCHEDULED REWARD GIVEN BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO

7. BLOCK DESIGN SUBTEST

SCHEDULED REWARD GIVEN BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1 trial 1	YES / NO	YES / NO	YES / NO
1 trial 2	YES / NO	YES / NO	YES / NO
2 trial 1	YES / NO	YES / NO	YES / NO
2 trial 2	YES / NO	YES / NO	YES / NO
3 trial 1	YES / NO	YES / NO	YES / NO
3 trial 2	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO

8. VOCABULARY SUBTEST

SCHEDULED REWARD GIVEN BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO

11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO
19	YES / NO	YES / NO	YES / NO
20	YES / NO	YES / NO	YES / NO
21	YES / NO	YES / NO	YES / NO
22	YES / NO	YES / NO	YES / NO
23	YES / NO	YES / NO	YES / NO
24	YES / NO	YES / NO	YES / NO
25	YES / NO	YES / NO	YES / NO
26	YES / NO	YES / NO	YES / NO
27	YES / NO	YES / NO	YES / NO
28	YES / NO	YES / NO	YES / NO
29	YES / NO	YES / NO	YES / NO
30	YES / NO	YES / NO	YES / NO

9. OBJECT ASSEMBLY SUBTEST

SCHEDULED REWARD GIVEN BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO

10. COMPREHENSION SUBTEST

SCHEDULED REWARD GIVEN BETWEEN SUBTESTS	YES	NO

ITEM	ITEM ADMINISTERED	REWARD SCHEDULED	REWARD GIVEN
1	YES / NO	YES / NO	YES / NO
2	YES / NO	YES / NO	YES / NO
3	YES / NO	YES / NO	YES / NO
4	YES / NO	YES / NO	YES / NO
5	YES / NO	YES / NO	YES / NO
6	YES / NO	YES / NO	YES / NO
7	YES / NO	YES / NO	YES / NO
8	YES / NO	YES / NO	YES / NO
9	YES / NO	YES / NO	YES / NO
10	YES / NO	YES / NO	YES / NO
11	YES / NO	YES / NO	YES / NO
12	YES / NO	YES / NO	YES / NO
13	YES / NO	YES / NO	YES / NO
14	YES / NO	YES / NO	YES / NO
15	YES / NO	YES / NO	YES / NO
16	YES / NO	YES / NO	YES / NO
17	YES / NO	YES / NO	YES / NO
18	YES / NO	YES / NO	YES / NO

STANDARD WISC-III ADMINISTRATION WAS MAINTAINED (EXCEPT FOR REWARD TREATMENT)	YES	NO

TOTAL REWARDS SCHEDULED PER TREATMENT PROCEDURE	
TOTAL REWARDS GIVEN PER RATER OBSERVATION	

Appendix L

Request for Permission to Reprint from DSM-IV and Associated Correspondence

The Graduate School and University Center of the City
University of New York
Educational Psychology Department
33 West 42 Street
New York, New York 10036-8099
212-642-2261

August 31,1998

Division of Publications and Marketing
American Psychiatric Association
1400 K Street, N.W.
Washington, DC 20005

Dear Division of Publications and Marketing:

I am requesting permission to reproduce an excerpt from DSM-IV for inclusion in my doctoral dissertation entitled "The Effects of Different Incentive Conditions on the WISC-III Performance of Conduct Disorder Adolescents". This dissertation is being completed under the supervision of Professor Philip A. Saigh, Head of the School Psychology Program in the Educational Psychology Department at The Graduate School and University Center of the City University of New York.

I have completed and am enclosing the APA's Request for Permission to Reprint from DSM-IV which you have sent me. I hope to cite selected text regarding the diagnostic criteria for Conduct Disorder on pages 90 to 91. I am requesting that the administrative fee of $50 be waived because my request is for educational use, and I do not expect to profit financially from the use.

I appreciate your considering this request and eagerly await your response.

Sincerely,

Edmond Fallon
Doctoral Student

American Psychiatric Association

1400 K Street, N. W.
Washington, D. C. 20005
Telephone 202.682.6000
Fax 202.682.6850
E-mail apa@psych.org
Internet: www.psych.org

September 28, 1998

Edmond Fallon
2923 Old Yorktown Rd.
Yorktown Heights, NY 10598

Dear Ms. Fallon,

I am responding to your recent request to **reprint the diagnostic criteria for Conduct Disorder,** from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.

Permission is **granted** under the following conditions:

- Permission is nonexclusive and limited to the single use specified in your letter;
- Use is limited to the English language only; and
- Permission must be requested for subsequent uses (including subsequent editions)

The APA's Board of Trustees set policy and guidelines for handling requests to reprint from DSM-IV several years ago. Authors/editors requesting permission to reprint material that will be used in books and plan to be sold are charged fees based on the amount of material they wish to reprint. Fees are not assessed for educational or training uses.

In all instances, the source and copyright status of the reprinted material must appear with the reproduced text. The following notice should be used:
Based on information from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Copyright 1994 American Psychiatric Association.

The correct bibliographic citation for DSM-IV is as follows:
American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC, American Psychiatric Association, 1994.

Sincerely,

Ronald E. McMillen /PA

Ronald E. McMillen
Director of Publications

Enclosure

References

Achenbach, T. M. (1993). Taxonomy and comorbidity of conduct problems: Evidence from empirically based approaches. Development and Psychopathology, 5, 51-64.

Amante, D. (1976). Visual-motor malformation, ethnicity, and social-class position. Journal of Special Education, 9, 247-249.

American Psychiatric Association. (1952). Diagnostic and statistical manual of mental disorders (1st ed.). Washington, DC: Author.

American Psychiatric Association. (1968). Diagnostic and statistical manual of mental disorders (2nd ed.). Washington, DC: Author.

American Psychiatric Association. (1968). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd ed. - revised). Washington, DC: Author.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author.

Ammons, R. B., & Ammons, C. H. (1962). The Quick Test (QT): Provisional manual. Psychological Reports, 11, 111-161.

Anastasi, A. (1997). Psychological testing (7th ed.). New York: Macmillan.

Barclay, R.A. (1990). Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment. New York: Guilford.

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, 12, 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)

Bradley-Johnson, S., Johnson, C. M., Shanahan, R., Rickert, V., & Tardona, D. (1984). Effects of token reinforcement on WISC performance of Black and White, low socioeconomic second graders. Behavioral Assessment, 6, 365-373.

Bradley-Johnson, S., Graham, D., & Johnson C. M., (1986). Token reinforcement on WISC-R performance for White, low socioeconomic, upper and lower elementary-school-age students. Journal of School Psychology, 24, 73-79.

Breuning, S. E., & Zella, W. F. (1978). Effects of individual incentives on norm-referenced IQ test performance of high school students in special education classes. Journal of School Psychology, 16, 220-226.

Busch, J. C., & Osborn, W. L. (1976). Significant vs. meaningful differences in the effects of tangible reinforcement on intelligence test achievement and reliability of TMR subjects. Psychology in the Schools, 13, 219-225.

Cairns, R. B., Cairns, B. D., & Neckerman, H. J. (1989). Early school dropout: Configurations and determinants. Child Development, 60. 1437-1452.

Cairns, R. B., Cairns, B. D., Neckerman, H. J., Gest, S. D., & Gariepy, J. L. (1988). Social networks and aggressive behavior: Peer support or peer rejection? Developmental Psychology, 24, 815-823.

Campbell, D. T., & Stanley, J. C. (1963). Quasi-experimental designs for research. Hopewell, NJ: Houghton Mifflin.

Cantwell, D. P. (1989). Conduct disorder. In H. I. Kaplan & B. J. Sadock (Eds.). Comprehensive textbook of psychiatry, (5th ed., pp. 1821-1828). Baltimore: Williams and Wilkens.

Clingman, J., & Fowler, R. L. (1975). The effects of contingent and noncontingent rewards on the I.Q. scores of children of above-average intelligence. Journal of Applied Behavior Analysis, 8, 90.

Cohen, J. (1988). Statistical power analysis for the behavior sciences. Hillsdale, NJ: Erlbaum Associates.

Cohen, L. (1970). The effects of material and non-material reinforcement upon performance of the WISC Block Design subtest by children of different social classes: A follow-up study. Psychology, 7, 41-47.

Cook, L. C. (1973). The effects of verbal and monetary feedback on the WISC scores of lower-SES Spanish American and lower- and middle-SES Anglo students. Dissertation Abstracts International, 34, 1693A-1694A. (University Microfilm No. 72-23,317)

Costello, E. J. (1989). Developments in child psychiatric epidemiology. Journal of the American Academy of Child and Adolescent Psychiatry, 28, 836-841.

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

Daugherty, T. K., & Quay, H. C. (1991). Response perseveration and delayed responding in childhood behavior disorders. Journal of Child Psychology and Psychiarty, 32, 453-461.

Daugherty, T.K., Quay, H. C. & Ramos, L. (1991). Response perseveration, inhibitory control, and central dopaminergic activity in childhood behavior disorders. Unpublished manuscript, University of Miami Coral Cables FL.

Devers, R., Bradley-Johnson, S., & Johnson, C. M. (1994). The effect of token reinforcement on WISC-R performance for fifth- through ninth-grade American Indians. Psychological Record, 44, 441-449.

Dietrich, C. (1976). Differential effects of task and reinforcement variables on the performance of three groups of behavior problem children. Journal of Abnormal Child Psychology, 4, 155-171.

Dishion, T.J., Patterson, G. R., & Stoolmiller, M. (1991). Family, school and behavioral antecedents to early adolescent involvement with antisocial peers. Developmental Psychology, 27,172-180.

Dodge, K. A., Cole, J. D., & Brakke, N. P. (1982). Behavior patterns of socially rejected and neglected preadolescents: The role of social approach and aggression. Journal of Abnormal Child Psychology, 10, 389-410.

Dodge, K. A., Price, J. M., Bachorowski, J., & Newman, J.P. (1990). Hostile attributional biases in severely aggressive adolescents. Journal of Abnormal Psychology, 99, 385-392.

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

Edlund, C. (1972). The effect on the behavior of children, as reflected in the IQ scores, when reinforced after each correct response. Journal of Applied Behavior Analysis, 5, 317-319.

Farrington, D. P. (1991). Childhood aggression and adult violence: Early precursors and later life outcomes. In D. J. Pepler & K. H. Rubin (Eds.). The development and treatment of childhood aggression (pp. 5-29). Hillsdale, NJ: Lawrence Erlbaum.

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

Fonseca, A. C., & Yule, W. (1995). Personality and antisocial behavior in children and adolescents: An inquiry into Eysenck's and Gray's theories. Journal of Abnormal Child Psychology, 23, 767-781.

Galbraith, G., Ott, J., & Johnson C. M. (1986). The effects of token reinforcement on WISC-R performance of low-socioeconomic Hispanic second graders. Behavior Assessment, 8, 191-194.

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.

Geffner, D. S., & Hochberg, I. (1976). The cognitive use of language by disadvantaged and non-disadvantaged children. Journal of Communication Disorders, 8, 31-42.

Gleuck, S., & Gleuck, E. (1968). Delinquents and nondelinquents in perspective. Cambridge, MA: Harvard University Press

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.

Golden, C. (1978). Stroop Color and Word Test. Woodale, IL: Stoetling & Company.

Gottfried, A. W. (1985). Measures of Socioeconomic status in child development research. Merrill-Palmer Quarterly, 31, 85-92.

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.

Grant, D.A., Berg, E. A. (1993). Wisconsin Card Sorting Test. Tampa: Psychological Assessment Resources.

Gray, J. A. (1982). The neuropsychology of anxiety: An enquiry into functions of the septo-hippocampal system. Oxford: Oxford University Press.

Gray, J. A. (1987). The psychology of fear and stress, (2nd ed.). Cambridge: Cambridge University Press.

Gray-Ray, D., & Ray, M. C. (1990). Effects of family structure on black male delinquency. Youth and Society, 22, 67-84.

Hammill, D. D. (1985). Detroit Test of Learning Ability - 2. Austin: PRO-ED.

Henggeler, S. W. (1989). Delinquency in adolescence. Newbury Park, CA: Sage.

Herjanic, B., & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. Journal of Abnormal Psychology, 10, 307-324.

Heusmann, L. R., Eron, L. D., Lefkowitz, M. M., & Walder, L. O. (1984). The stability of aggression over time and generations. Developmental Psychology, 20, 1120-1134.

Hinshaw, S. P., Lahey, B. B., & Hart, E. L. (1993). Issues of taxonomy and comorbidity in the development of conduct disorder. Development and Psychopathology, 5, 31-49.

Hodges, K., & Plow, J. (1990). Intellectual ability and achievement in psychiatrically hospitalized children with conduct, anxiety, and affective disorders. Journal of Consulting and Clinical Psychology, 5, 589-595.

Hollingshead, A. B. (1957). Two factor index of social position. Unpublished manuscript, Yale University Department of Sociology, New Haven, CT.

Hollingshead, A. B. (1975). Four factor index of social status. Unpublished manuscript, Yale University, Department of Sociology, New Haven, CT.

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.

Institute of Medicine. (1989). Research on children and adolescents with mental, behavioral, and developmental disorders. Washington, DC: National Academy Press.

Jackson, A. M., Farley, G. K., Zimet, S. G., & Gottman J. M. (1979). Optimizing the WISC-R test performance of low- and high-impulsive emotionally disturbed children. Journal of Learning Disabilities, 12, 622-625.

Jessor, R., & Jessor, S. L. (1977). Problem behavior and psychological development: A longitudinal study of youth. New York: Academoic Press.

Johnson, C. M., Bradley-Johnson, S., McCarthy, R. & Jamie, M. (1984). Token reinforcement during WISC-R administration: II. Effects on mildly retarded, Black students. Applied Research in Mental Retardation, 5, , 43-52.

Kalantari, M., Yule, W., & Gardner, F. (1990, Jan.). Oversensitivity to reward in preschool conduct disordered children. Paper presented at the annual meeting of the Society for Research in Child and Adolescent Psychopathology, Costa Mesa, CA.

Kazdin, A. E. (1985). Treatment of antisocial behavior in children and adolescents. Homewood, IL: Dorsey Press.

Kazdin, A. E. (1996). Conduct disorders in childhood and adolescence (2nd ed.). Thousand Oaks, CA: Sage.

Kazdin, A. E., Siegel, T. C., & Bass, D. (1990). Drawing upon clinical practice to inform research on child and adolescent psychotherapy: A survey of practitioners. Professional Psychology: Research and Practice, 21, 189-198.

Kendall, P. & Hammen, C. (1996). Abnormal Psychology. Boston, MA: Houghton Mifflin.

Kennedy, W. A., & Willcutt, H. C. (1964). Praise and blame as incentives. Psychological Bulletin, 61, ,323-332.

Kieffer, D. A., & Goh, D. S. (1981). The effect of individually contracted incentives on intelligence test performance of middle- and low-SES children. Journal of Clinical Psychology, 37, 175-179.

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

Ladd, G. S., Hart, C. H., & Price, J. M. (1990). Preschoolers' behavioral orientations and patterns of peer control: Predictive of peer status? In S. R. Asher & J. D. Cole (Eds.), Peer rejection in childhood. Cambridge: Cambridge University Press.

Last, J. M. (1983). A dictionary of epidemiology. New York: Oxford University Press.

Ledingham, J. E., & Schwartzman, A. E. (1984). A 3-year follow-up of aggressive and withdrawn behavior in childhood: Preliminary findings. Journal of Abnormal Child Psychology, 12, 157-168.

Levin, G. R., & Simmons, J. J. (1962a). Response to food and praise by emotionally disturbed boys. Psychological Reports, 11, 10.

Levin, G. R., & Simmons, J. J. (1962b). Response to food and praise by emotionally disturbed boys. Psychological Reports, 11, 539-546.

Loeber, R. (1990). Development and risk factors of juvenile antisocial behavior and delinquency. Clinical Psychology Review, 10, 1-14.

Loeber, R., Lahey, B. B., & Thomas, C. (1991). Diagnostic conundrum of oppositional defiant disorder and conduct disorder. Journal of Abnormal Psychology, 100, 337-352.

Loeber, R., & Schmaling, K. B. (1985). The utility of differentiating between mixed and pure forms of antisocial child behavior. Journal of Abnormal Child Psychology, 13, 315-335.

Loeber, R., & Stouthamer-Loeber, M. (1998). Development of juvenile aggression and violence. American Psychologist, 53, 242-259.

Lorge, I., & Thorndike, L. (1957). The Lorge-Thorndike Intelligence Tests. Boston: Houghton Mifflin.

Lyle, J. G., & Johnson, E. G. (1973). Analysis of WISC Coding: 3. Writing and copying speed, and motivation. Perceptual and Motor Skills, 36, 211-214.

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)

McCord, W., McCord, J., & Zola, I. K. (1959). Origins of crime. New York: Columbia University Press.

McGee, R., Williams, S., & Silva P., A. (1984). Behavioral and developmental characteristics of aggressive, hyperactive and hyperactive-aggressive boys. Journal of the American Academy of Child Psychiatry, 23, 270-290.

McGhee, R. L., & Short, R. J. (1991). The prevalence of social maladjustment among school-aged children. Psychology in the Schools, 28, 285-289.

Miller, D. C. (1970). Handbook of research design and social measurement. (2nd ed.). New York: David McKay.

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

Moffit, T.E. (1990). Juvenile delinquency and attention-deficit disorder: Developmental trajectories from age 3 to 13. Child Development, 61, 893-910.

Moffit, T.E. (1993a). Adolescence-limited and life-course persistent antisocial behavior: A developmental taxonomy. Psychological Review, 100, 674-701.

Moffit, T. E. (1993b). The neuropsychology of conduct disorder. Development and Psychopathology, 5, 135-151.

Morey, L. C., Skinner, H. A., & Blashfield, R. K. (1986). Trends in the classification of abnormal behavior. In A. R. Cimenero, K. S. Calhoun, & H. E. Adams (Eds.), Handbook of behavioral assessment, (2nd ed., pp. 47-78). New York: John Wiley & Sons.

National Association of School Psychologists. (1992). Principles for professional ethics. Silver Spring, MD: Author.

National Opinion Research Council. (1974). National data program for the social sciences, codebook for the Spring 1974, general social survey. Pittsfield, MA: Roper Public Opinion Research Center.

Offord, D. R., Boyle, M. H., & Racine, Y. A. (1991). The epidemiology of antisocial behavior. In D. J. Pepler & K.H. Rubin (Eds.), The development and treatment of childhood aggression (pp. 31-54). Hillsdale, NJ: Lawrence Erlbaum.

Otis, A. S. (1918). Otis Self-Administering Test. New York: World Book Company.

Otis, A. S. (1924). Otis Intelligence Scale Primary Examination. New York: World Book Company.

Otis, A. S. (1954). Otis Quick Scoring Mental Ability Test. New York: Harcourt-Brace-Jovanovich.

Otis, A. S., & Lennon, R. T. (1982). Otis-Lennon school ability test: Manual for administration and interpreting. (Elementary Form R). San Antonio, TX: Harcourt Brace Jovanovich.

Patterson, G.R. (1992). Developmental changes in antisocial behavior. In R. D. Peters, R. J. McMahon, & V. L. Quincy (Eds.). Aggression and violence throughout the lifespan (pp. 52-82). Newbury Park, CA: Sage.

Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. American Psychologist, 44, 329-335.

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). Antisocial boys. Eugene, OR: Castalia.

Peniston, E., & McLean, M. G. (1979). Parental influence on maladaptive behavior of mentally retarded children in the classroom. British Journal of Mental Subnormality, 25, 70-74.

Peterson, D. R., Quay, H. C., & Tiffany, T. L. (1961). Personality factors related to juvenile delinquency. Child Development, 32, 355-372.

Pollock, M. A. (1990). Incentive conditions and the selected WISC-R subtest performance of elementary school children (Doctoral dissertation, City University of New York. 1989). Dissertation Abstracts International, 51, 800.

Porteus, S. E. (1942). Qualitative performance in the Maze Test. Vineland, NJ: Smith.

Porteus, S. E. (1965). Porteus Maze Tests. Palo Alto, CA: Pacific Books.

Pullis, M. (1991). Practical considerations of excluding conduct disordered students: An empirical analysis. Behavior Disorders, 17, 9-22.

Quay, H. C. (1964a). Personality dimensions in delinquent males as inferred from the factor analysis of behavior ratings. Journal of Research in Crime and Delinquency, 1, 33-37.

Quay, H. C. (1964b). Dimensions of personality in delinquent boys as inferred from the factor analysis of case history data. Child Development, 35, 479-484.

Quay, H. C. (1986). Conduct disorders. In H. C. Quay & J. S. Werry (Eds.). Psychopathological disorders of childhood (3rd ed.). New York, NY: John Wiley.

Quay, H. C. (1988a). Attention deficit disorder and the behavioral inhibition system: The relevance of the neuropsychological theory of Jeffrey A. Gray. In L. M. Bloomingdale & J. Sergeant (Eds.). Attention deficit disorder: Criteria, cognition, intervention, (pp. 117-125). Oxford: Pergamon.

Quay, H. C. (1988b). The behavioral reward and inhibition systems in childhood behavior disorders. In L. M. Bloomingdale (Ed.). Attention deficit disorder (Vol. 3, pp. 176-186). Oxford: Pergamon.

Quay, H. C. (1988c). Reward, inhibition, and ADDH. Journal of the American Academy of Child and Adolescent Psychiatry, 27, 262-263.

Quay, H. C., & Peterson, D. R. (1967). Manual for the Behavior Problem Checklist. Champaign, IL: University of Illinois.

Quay, H. C., & Peterson, D. R. (1987). Manual for the Revised Behavior Problem Checklist. Coral Gables, FL: Authors.

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.

Raven, J. C. (1960). Guide to the Standard Progressive Matrices. London: Lewis.

Raven, J. C. (1965). Guide to the Standard Progressive Matrices. London: Lewis.

Reich, W., Leacock, N., & Shanfeld, K. (1995). Diagnostic interview for children and adolescents DICA-R-P child, adolescent, and parent versions, DSM-III-R and DSM-IV versions. St. Louis, MO: Division of Child Psychiatry, Washington University.

Reitsma-Street, M., Offord, D. R., & Finch, T. (1985). Pairs of same-sexed siblings discordant for antisocial behavior. British Journal of Psychiatry, 146, 415-423.

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

Richman, L. C., & Lindgrin, S. D. (1981). Verbal mediation deficits: Relation to behavior and achievement in children. Journal of Abnormal Psychology, 90, 99-104.

Robins, L. N. (1978). Sturdy childhood predictors of adult antisocial behavior: Replications from longitudinal studies. Psychological Medicine, 8, 611-622.

Robins, L. N. (1981). Epidemiological approaches to natural history research: Antisocial disorders in children. Journal of the American Academy of Child Psychiatry, 20, 566-680.

Robins, L. N. (1991). Conduct disorder. Journal of Child Psychology and Psychiatry, 32, 193-212.

Rutter, M. (1967). Children's behavior questionnaire for completion by teachers. Journal of Child Psychology and Psychiatry, 8, 1-11.

Rutter, M., & Giller, H. (1983). Juvenile delinquency: Trends and perspectives. New York: Penguin.

Saigh, P. A. (1981a). 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. (1981b). The validity of the WISC-R examiner verbal praise procedure as a concurrent predictor of intellectually superior students. Journal of Clinical Psychology, 37, 647-649.

Saigh, P. A. (1982, August). Incentive conditions and the WISC-R performance of institutionalized conduct disordered adolescent females. Paper presented at the meeting of the American Psychological Association, Washington, DC.

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, 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.

Sattler, J. M. (1992). Assessment of children (3rd ed.). San Diego, CA: Author.

Shapiro, S. K., & Hynd, G. W. (1993). Psychological basis of conduct disorder. School Psychology Review, 22, 386-402.

Shapiro, S. K., Quay, H. C., Hogan, A. E., & Schwartz, K. P. (1988). Response perseveration and delayed responding in undersocialized conduct disorder. Journal of Abnormal Psychology, 97, 371-373.

Short, R. J., & Shapiro, S. K. (1993). Conduct disorders: A framework for understanding and intervention in schools and communities. School Psychology Review, 22, 362-375.

Silverstein, A. B. (1967). Validity of WISC short form at three age levels. Journal of Consulting Psychology, 31, 635-636.

Slosson, R. L. (1963). Slosson Intelligence Test for Children and Adults. New York: Slosson Educational Publications.

Slosson, R. L. (1975). Slosson Intelligence Test for Children. New York: Psychological Corporation.

Slotkin, J., Forehand, R., Fauber, R., McCombs, A., & Long, N. (1988). Parent-completed and adolescent-completed CDIs: Relationship to adolescent social and cognitive functioning. Journal of Abnormal Child Psychology, 16, 207-217.

Smeets, P. M., & Striefel, S. (1975). The effects of different reinforcement conditions on the test performance of multihandicapped deaf children. Journal of Applied Behavior Analysis, 8, 83-89.

Spivack, G., Pratt, J. J., & Shure, M. B. (1976). The problem solving approach to adjustment. San Francisco: Jossey-Bass.

Stewart, D. J. (1972). Effects of social reinforcement on dependency and aggressive responses of psychopathic, neurotic and sub-cultural delinquents. Journal of Abnormal Psychology, 79, 76-83.

Suh, M., & Carlson, R. (1977). Childhood behavioral disorders: A family typology. Psychiatric Journal of the University of Ottawa, 2, 84-88.

Sullivan, E. T., Clark, W. W., & Tiegs, E. W. (1957). California Test of Mental Maturity (1975 ed.). Monterey, CA: California Test Bureau.

Sweet, R., & Ringness, T. (1971). Variation in the intelligence test performance of referred boys of differing racial and socioeconomic backgrounds as a function of feedback or monetary reinforcement. Journal of School Psychology, 9, 399-409.

Tellegen, A., & Briggs, P.F. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. Journal of Consulting Psychology, 31, 499-506.

Terman, L. (1916). The measurement of intelligence. Boston: Houghton-Mifflin.

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

Terman, L. M., & Merrill, M. A. (1960). Stanford-Binet Intelligence Scale. Boston: Houghton-Mifflin.

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.

Tharp, R. G., & Wetzel, R. S. (1971). Behavior modification in the natural environment. New York: Academic Press.

Thorndike, E. (1924). Measurement of intelligence. Psychological Review, 31, 219-252.

Tiber, N., & Kennedy, W. A. (1964). The effects of incentives on the intelligence test performance of different social groups. Journal of Consulting Psychology, 28,187-189.

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

Warren, L. W., Meeker, N., & Eells, K. (1960). Social class in America: A manual for procedures for the measurement of social status. New York: Harper & Row.

Wechsler, D. (1949). Wecheler intelligence Scale for Children. New York, NY: The Psychological Corporation.

Wechsler, D. (1967). Wechsler preschool and primary scale in intelligence. New York, NY: The Psychological Corporation.

Wechsler, D. (1974). Wecheler intelligence Scale for Children - Revised. San Antinio, TX: The Psychological Corporation.

Wechsler, D. (1991). Wecheler intelligence Scale for Children - Third Edition. San Antinio, TX: The Psychological Corporation.

West, M. O., & Prinz, R. J. (1987). Parental alcoholism and childhood psychopathology. Psychological Bulletin, 102, 203-218.

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

Willis, J., & Shabata, B. (1978). A comparison of tangible reinforcement and feedback effects on the WPPSI IQ scores of nursery school children. Education and Treatment of Children, 1, 31-45.

Wilson, H. (1980). Parent supervision: A neglected aspect of delinquency. British Journal of Criminology, 20, 203-235.

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.

Wolfgang, M.E., Figlio, R. & Sellin, T. (1972). Delinquency in the birth cohort. Chicago: University of Chicago Press.

World Health Organization. (1992). International classification of diseases (ICD-10). Geneva: Author.

Young, R. M., Bradley-Johnson, S., & Johnson C. M. (1981). Immediate and delayed reinforcement on WISC-R performance of mentally retarded students. Applied Research in Mental Retardation, 3, 13-20.

Zoccolillo, M. (1992). Co-occurrence of conduct disorder and its adult outcomes with depressive and anxiety disorders: A review. Journal of the American Academy of Child and Adolescent Psychiatry, 31,547-556.

Zoccolillo, M. (1993). Gender and the development of conduct disorder. Development and Psychopathology, 5, 65-78.