72-10,165

KUBANY, Edward S., 1938-
THE EFFECTS OF INCENTIVES ON THE TEST
PERFORMANCE OF HAWAIIANS AND CAUCASIANS.

University of Hawaii, Ph.D., 1971
Psychology, clinical

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THE EFFECTS OF INCENTIVES ON THE TEST PERFORMANCE OF HAWAIIANS AND CAUCASIANS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN PSYCHOLOGY

SEPTEMBER 1971

By

Edward S. Kubany

Dissertation Committee:

Ronald Gallimore, Chairman
Karl A. Minke
Arthur W. Staats
Roland G. Tharp
Robert Earickson

ACKNOWLEDGMENTS

There are a great many people in the State of Hawaii Department of Education to whom I owe a huge debt of thanks for allowing me to conduct my research in their schools, for going out of their way to make it possible for my research to go as smoothly as it did, and in general for making it a real pleasure for me to do my doctoral research in the schools of the State of Hawaii. I would like to acknowledge, in particular, the following persons in the Department of Education: James Aki, Fred Cachola, Phil Iha, Ronald Johnson, Bernice Loo, Domingo Los Banos, Robert Main, Albert Miyasato, Robert Ogata, Cyril Okamoto, Lynn Schutte, and Marion Selig.

I would also like to gratefully acknowledge the assistance of Lola Jordan who collected the Salt Lake City data.

ABSTRACT

The central question posed by the present research was, can the introduction of incentives into the test situation have a bearing on the level of test performance? A review of the literature revealed conflicting results.

Traditional theories of motivation have assumed that test motivation reflects a more general, intrapsychic "need" state which is not readily alterable; hence the belief that attempts to manipulate test motivation cannot result in improved test performance. In contrast, theories of motivation based on learning principles place much greater emphasis on situational determinants of "motivated behavior" while interpreting "intrinsic motivation" in terms of previous reinforcement history.

Three experiments were conducted in order to test the hypothesis that the introduction of incentives will result in test score increments. Experiments One and Two also tested the hypothesis that indigenous Hawaiians would be more influenced by group incentives and that Caucasians would be more affected by individual incentives.

Experiment One--considered a pilot study--revealed no evidence that the promise of reward can alter multiple-choice test math scores in a naturalistic school setting. Alternate explanations for the nonsignificant results were offered. However, the most important conclusion drawn from Experiment

One is that standardized multiple-choice achievement tests provide virtually no information about the achievement levels of many Hawaiians, and discontinuance of their use in favor of free-response tests is recommended.

In Experiment Two, which was conducted in a more laboratory-oriented atmosphere than Experiment One, sixth grade students were not only promised money but actually given money immediately contingent upon correct test responses. Relative to no-incentive conditions, subjects receiving incentives obtained significantly higher test scores and spent significantly longer periods of time working on their tests. There was also some evidence that the greatest "perseverance" was demonstrated by the Hawaiian subjects in the group-incentive condition.

Experiment Three was conducted in the naturalistic atmosphere of a youth correctional facility. Results showed that not only did test scores go up when incentives were introduced, but that when incentives were withdrawn, scores returned to near baseline levels. These results were interpreted as calling into question the validity of test results obtained from "motivationally deficient" youth and give rise to the recommendation that--with such populations--tests should be administered under both standard and incentive conditions.

General recommendations for future research were discussed, and an expanded diagnostic system, including the

classification "motivational disability" was proposed. Staats' developmental-learning conception of motivation (the A-R-D system) was used as the theoretical framework for the expanded classification system, and on the basis of Staats' A-R-D system a new model for child assessment in clinical psychology was derived and described.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

ABSTRACT iv

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER I MOTIVATION AND TEST PERFORMANCE 1

Achievement Motivation (as a Personality "Trait") and Test Performance 6
Incentives and Test Performance 7

CHAPTER II A BEHAVIORAL ANALYSIS OF ACHIEVEMENT MOTIVATION 11

CHAPTER III SOCIOECONOMIC AND ETHNIC DIFFERENCES IN ACHIEVEMENT MOTIVATION 13

Differential Effects of Extrinsic Incentives 13
Achievement Motivation Among Indigenous Hawaiians 16

CHAPTER IV RECAPITULATION 18

General Predictions 19
Implications of Positive Findings 20
Use of Money as an Incentive 21

CHAPTER V EXPERIMENT ONE: A PILOT STUDY 23

Introduction 23
Method 24
Subjects 24
Test Used 25
Experimental Design 25
General Procedures 26
Instructions 27
Results 29
Discussion 31
Instructional (Discriminative Stimulus) Control versus Reinforcement Control 31
Multiple-Choice Versus Free-Response Tests 32

	Task Difficulty.	33
	The Use of Standardized Multiple-Choice Achievement Tests in Hawaii	34
CHAPTER VI	EXPERIMENT TWO	38
	Method	39
	Subjects	39
	Test Used.	39
	Experimental Design.	40
	General Procedures	40
	Instructions	42
	Results.	45
	Discussion	56
CHAPTER VII	EXPERIMENT THREE	60
	Method	60
	Subjects	60
	Experimental Design.	60
	General Procedures	61
	Instructions	61
	Results.	62
	Discussion	66
	"Perseverance" Under Intrinsic and Extrinsic Test Conditions as a Measure of Motivation.	69
	Feedback--Reinforcement Confounding.	70
CHAPTER VIII	GENERAL DISCUSSION	71
	Recommendations for Future Research. .	71
	Differential Diagnosis and Prescription in Education	72
	Learning Disability--No Motivational Disability (LD-No MD).	73
	Learning Disability--Motivational Disability (LD-MD)	74
	No Learning Disability--Motivational Disability (No LD-MD).	77
	No Learning Disability--No Motivational Disability (No LD-No MD)	79
	Staats' A-R-D System	81
	"Motivational Disabilities" as A-R-D System Defects	85

Learning Disabilities as Instrumental Behavior Deficits. 87
Interaction of "Motivational Disability" (A-R-D Defect) and Learning Disability (Instrumental Behavior Deficit) 88
No "Motivational Disability" -- No Learning Disability. 91
The A-R-D Function of Language 93
A Clinical Psychology Model for Child Assessment 95

REFERENCES 103

LIST OF TABLES

1 Means and Standard Deviations of STEP Math Difference-Scores. 30

2 3X2 Analysis of Variance of STEP Math Difference-Scores 30

3 2X2 Analysis of Variance of STEP Math Difference-Scores (Incentive Conditions Collapsed). 32

4 Analysis of Variance of Sullivan Math Raw Difference-Scores 47

5 Analysis of Variance of Sullivan Math Book-Level Difference-Scores. 49

6 Analysis of Variance of Sullivan Math Time Difference-Scores. 50

7 Analysis of Covariance of Sullivan Math Time Difference-Scores. 53

8 Analysis of Variance of Sullivan Math Raw Scores (Repeated Measures). 65

9 Proposed Diagnostic System. 73

10 Testing Conditions in Proposed Child Assessment Model. 99

LIST OF FIGURES

1 Mean Raw Difference-Scores Obtained by Hawaiians and Caucasians on Sullivan Math Placement Test Under No-Incentive (Control), Individual-Incentive, and Group-Incentive Conditions. 46

2 Mean Book-Level Difference-Scores Obtained by Hawaiians and Caucasians on Sullivan Math Placement Test Under No-Incentive (Control), Individual-Incentive, and Group-Incentive Conditions. 48

3 Mean Time Difference-Scores Obtained by Hawaiians and Caucasians on Sullivan Math Placement Test Under No-Incentive (Control), Individual-Incentive, and Group-Incentive Conditions. 51

4 Weighted Mean Time Difference-Scores Obtained by Hawaiians and Caucasians on Sullivan Math Placement Test Under No-Incentive (Control), Individual-Incentive and Group-Incentive Conditions (Pretest Scores Held Constant) . . . 54

5 Mean Raw Scores Obtained on Sullivan Math Placement Test by 14 Subjects Given Repeated Measures Under No-Incentive Baseline Conditions, Individual-Incentive Conditions, and No-Incentive Return-to-Baseline Conditions 64

6 Hypothetical Test Profile Obtained Under Standard and Incentive Conditions of Child Diagnosed as Learning Disability--No "Motivational Disability" 75

7 Hypothetical Test Profile Obtained Under Standard and Incentive Conditions of Child Diagnosed as Learning Disability--"Motivational Disability". 76

8 Hypothetical Test Profile Obtained Under Standard and Incentive Conditions of Child Diagnosed as No Learning Disability--"Motivational Disability" 78

9 Hypothetical Test Profiles Obtained Under Standard and Incentive Conditions of Children Diagnosed as No Learning Disability--No "Motivational Disability". 80

CHAPTER I

MOTIVATION AND TEST PERFORMANCE

Since the days of Tolman and Hull, psychologists have been concerned with the relationship between learning and performance. The animal learning literature is replete with studies designed to lay to rest the theoretical issue of whether or not, or to what extent, incentives are necessary in order for learning to occur (Tolman and Honzik, 1930; Thistlewaite, 1951; Bandura, 1965). A related practical issue concerns the role of motivation in human test performance. When a person is given an examination, what, if any, incentives are necessary to elicit what he has "learned"? Addressing this issue, Guilford (1967) stated recently that,

> When we administer aptitude tests, we ordinarily exhort examinees to do their best, and this stimulus, together with a general cultural encouragement to do well if not to excel in what they do, is depended upon to a large extent to help ensure relatively high and uniform levels of effort. Examinees from other cultures are at disadvantages because motivation is not sufficiently aroused or not sufficiently pointed toward making a good score, particularly in speed tests. Even within our own culture, we do not know to what extent different levels of motivation prevail during testing and what effect these differences may have upon scores. (P. 406)

Many in the testing field have assumed that test motivation is primarily a function of some inferred internal state, and that the "predisposition" to strive is

not readily alterable. From this view, extreme or "artificial" attempts to increase test motivation would be contraindicated at worst--contaminating valuable information about the subject's psychodynamics--or a waste of time at best. The test-taker's "motivational level" reflects an intrapsychic state which is recalcitrant to immediate change.

In sharp contrast with the above viewpoint, many behaviorists--noting the situational-specificity of much behavior--have emphasized the role of environmental factors which affect performance (Mischel, 1969; Klinger, 1967; Greenspoon & Gerston, 1967). While various theorists (e.g., Freud, Murray) have cherished the notion that much motivated behavior expresses some need acquired in the course of early socialization, behaviorally oriented clinicians--operating under the rubric of learning principles--have demonstrated repeatedly that situational reinforcement contingencies are often the dominant forces that shape the configuration of present behavior. For example, retarded, autistic and delinquent children thought to have profound motivational deficits have been shown to generate high "motivation" under appropriate stimulus (reinforcement) conditions (e.g., Birnbrauer, Wolf, Kidder and Tague, 1965; Lovaas, 1967; Cohen, 1968).

While intrinsic achievement motivation, as an internal state may take years to develop, achievement oriented

behavior (noteworthy by its persistance and high frequency of occurrence) has been shown to have a great deal of situational specificity, elicited or suppressed by environmental constraints.

Such developments may have implications for the process of assessment. Psychological tests have typically been used for purposes of diagnostic labeling and placement, but have seldom been used for specifying concrete training or treatment procedures. Diagnostic summaries rarely go beyond general recommendations for "psychotherapy," "more individual attention," or at best, recommendations about the content or level of remediation materials--answering the question "what," but not "how." Such current usage is probably due in large part to the domination of the testing movement by the psychodynamic psychotherapies and the presumption that motivational "deficits" cannot be readily altered. And because psychodynamic approaches are more concerned with making inferences about internal mechanisms than statements about behavior, behavior modification procedures tend to be unrelated to the information obtained from psychological tests.

Interestingly, individual differences in motivation have been largely ignored in the standardization of tests such as the Stanford-Binet and Wechsler Intelligence Scales, and almost all other tests for that matter. In the establishment of test norms, standardization has usually

referred to uniformity of procedure in administering and scoring a test (Anastasi, 1968), but little or no attempt has been made to obtain uniformity in the test taker's psychological state or level of motivation. For example, Terman and Merrill (1960) state (regarding the Stanford-Binet) that,

> To elicit the subject's best efforts and maintain both high motivation and optimal performance level through the testing session are the sine qua non of good testing, but the means by which these ends are accomplished are so varied as to defy specific formulation. (Pp. 50-51)

Terman and Merrill relegate the responsibility for obtaining such motivation to the "interpersonal know-how" of the examiner. To the extent that "high" or "low" motivation affects a test-taker's score, error is introduced, and a given score thus becomes less directly comparable to the test's norms--due to an ability/motivation confounding.

Inability to hold constant or measure the effects of motivation in test-taking can have potentially serious implications. Kubany and Sloggett (1971) report such a case of a 9th grade boy who was about to be erroneously labeled "learning disability" on the basis of his scores on a conventionally administered learning disability test battery. According to the test results, the boy, Michael, was functioning several years below his grade level, and plans were being made to refer him to a special learning disability program. At about the same time, Michael had

also been referred to the school counselor because of disruptive classroom behavior, and as a result the second author observed him in the classroom on several occasions --recording the frequency of his disruptive behavior. Although these observations confirmed that Michael was indeed frequently disruptive, he was also observed to attend to movies for long periods of time and to draw elaborately detailed pictures of sports cars. Such observations tended to contradict the learning disability label tentatively ascribed. Consequently, it was decided to readminister the learning disability test battery--but this time, Michael was given two cents for each correct response. The results of the readministered test battery indicated Michael to be at grade level or above on all three subtests. Had it not been for fortuitous circumstances, Michael would have been erroneously stigmatized with an "LD" label and referred to a learning disability classroom. A disturbing question is how many other children have been "innocently" mislabeled for similar reasons.

This illustration indicates the potential usefulness of distinguishing between a learning disability and a "motivational disability." A child for whom academic tasks have low intrinsic interest would not need the same kind of training program as a child who lacks some ability to process or assimilate sensory input. Ideally, an

aptitude or ability test might yield both an aptitude or ability score and a "motivation" score.

Achievement Motivation (as a Personality "Trait") and Test Performance

Test-taking motivation can be considered as one operation of achievement motivation, which has been defined as the intrinsic disposition to compete successfully against a standard of excellence (McClelland, Atkinson, Clark, and Lowell, 1953) and in similar terms as the "capacity for satisfaction" in the attainment of success (Atkinson, 1966). For individuals high in need achievement (n Ach),[1] the experience of success or achievement is presumably accompanied by a self-reinforcing sense of accomplishment. For example, Winterbottom (1953) found that children who were strong in achievement motive were rated by teachers as deriving more pleasure from success than children who were weak in achievement motive.

Thus one might expect a person to try to do well on a test or task to the extent that he gains satisfaction from getting a good score. Atkinson and Feather (1966) cite several studies in which high n Ach subjects showed greater task persistence than low n Ach subjects. In one of these studies (Atkinson and Litwin, 1960) which utilized

---

[1]Traditionally, n Ach has been measured in terms of the amount of achievement imagery evoked by modified TAT cards (McClelland, Clark, Roby, & Atkinson, 1949).

students in a college psychology course, high n Ach students spent more time working on their final exam and had higher final exam scores than students low in n Ach. There is also some evidence that in experimental situations, high n Ach subjects perform superior to low n Ach subjects on anagram tasks (Lowell, 1950) and simple arithmetic problems (Wendt, 1954).

Finally, Rosen (1956) found that high school students with high n Ach scores were significantly more likely than low n Ach students to achieve grades of at least "B."

Incentives and Test Performance

If a person doesn't care how well he does on a test, why should he try to do well? What would be his pay-off --especially if he expects to do poorly? One way to provide a pay-off might be to offer and deliver meaningful incentives in return for good performance or correct responses. Before proceeding, however, the meaning of the term incentive will be clarified. In the present paper the term incentive will be used to refer to an extrinsic stimulus (such as food, money, praise, etc.) which may possess both discriminative and reinforcing properties. The term incentive is being considered essentially synonomous with the term reward, but is being preferred here because it also carries the motivational implication that as a stimulus, an incentive may not only strengthen a response but also instigate or intensify a

response--depending upon its temporal contiguity to the response (Dinsmoor, 1950). In this regard, it may be important to note that the incentive operations in the various studies referred to below were not all equivalent. In some of these studies, incentives were both promised and delivered during the test session (Klugman, 1944). In others, incentives were promised but not actually delivered during the test session (Staats, Minke, Goodwin, & Landeen, 1967). And in still others, the response-incentive contingency was not specified to the test-taker but incentives were actually delivered during the test session (Tiber and Kennedy, 1964). That the promise of an incentive (discriminative stimulus) and the delivery of an incentive (reinforcement) may have different motivational properties will be discussed later.

Several studies have shown that extrinsic or situational variables can have a significant impact upon aptitude or ability test performance. Test scores have been influenced positively by praise (Bornstein, 1967), examiner warm (Exner, 1966), money (Staats, Minke, Goodwin, and Landeen, 1967) and the possibility of promotion (Burt and Williams, 1962). The effects found by Staats et al. were incidental to the major purpose of their study but are of interest here. Using the California Achievement Test and the Lorge-Throndike Intelligence Test as measures of reading improvement, the authors observed that many of

their subjects appeared to be answering some of the tests randomly. In an attempt to salvage some of their test data, posttest subjects were also given alternate forms of the above tests under incentive conditions--two cents offered for each correct answer and one cent lost for each incorrect answer. Subjects showed significant advances on both tests under the incentive conditions, and the improvement occurred even though fewer items, on the average, were answered under the incentive conditions.

More generally, numerous laboratory investigations have shown clearly that situational variables can influence the magnitude of achievement oriented behavior (Atkinson and O'Conner, 1966; Kubany, Gallimore, and Buell, 1970) and the expression of fantasy n Ach (Klinger, 1967; Douvan, 1956; McClelland, Clark, Roby, and Atkinson, 1949).

The evidence regarding incentive effects has not been all positive, however, as several investigators have found no significant incentive effects (Tiber and Kennedy, 1964; Benton, 1936; Ferguson, 1937; Klugman, 1944). For example, Tiber and Kennedy found no significant differences in Stanford-Binet performance between conditions of praise, reproof, candy reward, and no particular incentive--despite using a large sample. In addition, two of the principal textbooks on psychological testing make either depreciative or only fleeting reference to the role of incentives in test taking. In her 588 page text, Anastasi (1968) reserves

only two pages to a discussion of motivation, while Cronbach (1960) states categorically that,

> The test score is not readily altered by simple incentives. There have been many attempts to raise test performance by prizes, pep talks, and monetary payments for increases in score. Almost invariably, such attempts fail to produce appreciable improvement on ability tests over the scores earned under the regular conditions of administration. (Pp. 52-53)

CHAPTER II

A BEHAVIORAL ANALYSIS OF ACHIEVEMENT MOTIVATION

That there is a link between incentives and intrinsic achievement motivation may be demonstrated when the concept of intrinsic motivation is reinterpreted in terms of learning principles. According to Staats (1963) motivational terms such as needs and motives simply involve descriptions of different people in the strength and type of reinforcers that are effective for them. Note, for example, Winterbottom's (1953) finding that children strong in achievement motive derive more pleasure from success than children weak in achievement motive.

A "sense of satisfaction" or "pride in accomplishment" is equivalent to the learning theory conception of behavior that has acquired conditioned or secondary reinforcement value. Staats points out that doing something difficult is not inherently reinforcing (cf. Azrin, 1961; Hull, 1943), and that unless reinforcement for such behavior occurs, one would expect an organism to escape from hard work.

> The same is true of accomplishment, attaining high standards, excelling oneself, rivaling and surpassing others, and so on. These consequences are not by themselves positive reinforcers. Prior to the appropriate training, there is no reason to suppose that surpassing others, matching standards, and so on, will be reinforcing because they have in the past been

> paired with positive reinforcers. Children have to be trained to find winning reinforcing, to hold standards, and so forth. Anyone who has attempted to instruct small children in group games finds immediately that those aspects of "competitive" events that are reinforcers for adults may be quite neutral stimuli for children. (Staats, 1963, p. 294)

Staats adds,

> A child raised in a situation where middle-class achievement is not reinforced will not work for these consequences. The individual who has been deprived of a history where achievements have come to be reinforcing is frequently described as lazy, stupid, constitutionally inferior, and so on. Actually, however, this individual would not aspire to middle-class achievements--would not find them reinforcing--because he had not been so trained. (1963, p. 296)

Once again, the "capacity for satisfaction" to which Atkinson (1966) refers is conceived in learning terms as behavior which has acquired conditioned reinforcement value through previous pairings with other reinforcers. Further, reinforcing stimuli--as discriminitive stimuli--come to exert control over behaviors which are appropriate for obtaining them. Thus, after repeated pairings of success with outside or extrinsic reinforcement (praise, material incentives, privileges), "striving for success" behaviors may be expected to continue for some time in the absence of extrinsic reinforcement. Such behaviors are commonly referred to as being intrinsically motivated.

CHAPTER III

SOCIOECONOMIC AND ETHNIC DIFFERENCES IN ACHIEVEMENT MOTIVATION

There is evidence that social groups differ in the extent to which they place emphasis upon (reinforce) achievement striving. Using responses to a projective test as a measure of achievement motivation, several investigators have reported differences between ethnic and socioeconomic groups (Mingione, 1965; Rosen, 1956; Sloggett, Gallimore, and Kubany, 1970). For example, Rosen reported that high and middle class students wrote significantly more achievement themes than lower class students. In addition, on a measure of value orientation, Rosen found that the upper and middle class subjects were more likely than the lower class subjects to possess values which are likely to facilitate achievement and social mobility. Lastly, the upper class subjects also had higher grades than the lower class subjects.

Differential Effects of Incentives

Because of their different reinforcement histories, certain ethnic and socioeconomic groups--those lower in achievement motivation--may be more dependent on incentives than others. That is, certain groups may be less motivated than others in the absence of incentives. During an

aptitude test, for example, middle class Caucasians, high in intrinsic motivation, may perform near their motivational ceiling without incentives; as a result, incentives might not influence appreciably their test scores.

There is some evidence which bears on this hypothesis. Haggard (1954) offered high and low status subjects free theatre passes "if they did their best" on the Stanford-Binet and found that the procedure positively influenced the scores of the low status subjects, but had no influence on the scores of the high status subjects.

Klugman (1944) promised and gave pennies to Caucasian and Black subjects for correct solutions on the Stanford-Binet. Although the mean score for all subjects was not significantly higher under money conditions than under praise conditions, "there is substantial superiority of the scores of colored children who were tested with the money incentive. In addition, although there is practically no difference between the colored and white children when money is the incentive, the white show substantial superiority when [only] praise is the incentive." (Klugman, 1944, p. 268) (Incidentally, Mingione (1965) reported that low socioeconomic Caucasians have significantly higher n Ach scores than low socioeconomic Blacks.)

Douvan (1956) gave middle class and working class subjects a series of tasks after telling them either that

(1) good performance would bring a $10 prize, or (2) there would be no such prize. After the tasks, all subjects were told that they had failed and then were immediately administered the TAT measure of n Ach. The middle class subjects obtained similar and relatively high n Ach scores under both conditions (money/no money). In contrast, the working class subjects only had high n Ach scores under the condition of the $10 prize offer. Douvan concluded that,

> More autonomous and generalized success strivings characterize members of the middle class, while achievement motivation of working class individuals is more highly dependent on the reward loading of the task situation. (Douvan, 1956, p. 222)

In a study mentioned earlier, Tiber and Kennedy (1964) Stanford-Binet tested middle-class and lower-class Caucasian children as well as lower-class Black children under several incentive conditions (praise, reproof, candy, control). Contrary to the findings of Douvan (1956) and Haggard (1954), there were no significant differences in means between groups and no interactions between incentive operations and social class (in spite of rather large samples). Weaknesses in the experimental design, however, could have contributed to the nonsignificant results obtained. Not only were the subjects not informed of any incentive/performance contingency, but incentives were administered at the end of each subtest--regardless of the level of the subject's performance. Cooperation and

attention to the examiner may have been shaped up, but the test responses themselves were rewarded noncontingently--regardless of correctness--on a fixed ratio basis. Furthermore, the between groups analysis utilized by Tiber and Kennedy was less efficient than a repeated measures design would have been; that is, large individual differences in intelligence undoubtedly produced a very large error term.

Achievement Motivation Among Indigenous Hawaiians

There is evidence that indigenous Hawaiians, as a group, are seldom concerned with the pursuit of success for the purely personal satisfactions involved (Gallimore and Howard, 1968; Gallimore, 1969; Sloggett, Gallimore, and Kubany, 1970; Sloggett, 1969, 1970). Hawaiians write extremely few achievement themes to n Ach picture stimuli, and as a group have the poorest academic records in the State. Hawaiians apparently derive little personal pleasure from competing successfully against others, and in fact, avoid individual competition. According to Gallimore and Howard, Hawaiians are motivated primarily by peer approval, affiliation, and avoidance of disapproval, i.e., extrinsically! In a classroom field experiment, Sloggett (1969) found that low achieving Hawaiian adolescents preferred working in teams instead of as individuals, and that they showed marked preferences for rewards that could be shared. Anecdotally, she also observed that Hawaiian students would refuse to accept material rewards (e.g., cokes or

candy) for high grades or successful competition--unless the rewards could be shared with their friends.

Interestingly, Sloggett et al. (1970) found that high achieving Hawaiians, as a group, did not have higher need achievement scores than low achieving Hawaiians, and that their need achievement scores were unrelated to measures of intelligence, reading achievement and math achievement. Explaining these surprising findings, Sloggett et al. state that,

> The conclusion that Hawaiian males are poor scholars because they lack n Ach is a "deficiency" explanation (Gallimore, 1969); at best, it explains why Hawaiian children do not behave as middle-class Caucasians do, but it does not explain what motivates those among the Hawaiian children who do achieve. Thus the issue becomes what motive, if not n Ach, is associated with achievement for Hawaiian males. One preliminary study (Gallimore, 1969) indicates that for Hawaiian high school boys, fantasy n Affiliation and not n Ach, is correlated with achievement-oriented behavior in a risk taking task. This finding is consistent with the ethnographic data (Gallimore and Howard, 1968) which clearly reflected the importance attached by Hawaiians to affiliative rewards; thus, it is possible that among Hawaiians, both HA and LA, those individuals whose preferences for affiliative reinforcement are particularly strong may be motivated to work especially hard at achievement tasks in order to gain the approval of others. (Sloggett, et al., 1970, p. 9)

Since affiliation and peer approval are apparently reinforcing stimuli for Hawaiians, and since competition may even have aversive properties, one might suspect that social or group incentives would have greater reinforcing and discriminative stimulus value for Hawaiians than would individual incentives.

CHAPTER IV

RECAPITULATION

The preceding narrative has been guided by the assumption that achievement motivation--as a personality construct--is acquired on the basis of previous experiences in which "striving to do well" and "success" behaviors have been rewarded; "work" is inherently aversive and only becomes "its own reward" after being repeatedly paired with positive reinforcers.

In view of great individual variations--with respect to previous history of reinforcement for achievement-oriented behavior in general, and test-taking behavior in particular--motivational factors may occasionally affect the level of test performance to the extent that what a person "knows" or has learned becomes obscure. Thus one might predict that introducing incentives into the test situation will result in test score increments over scores obtained under standard testing conditions--especially for persons with a learning history devoid of reinforcement for doing well. Although some of the evidence reviewed indicates that the introduction of incentives can have a salutory effect on test performance, evidence from other sources is equivocal, and there seems to be a pervasive *a priori* acceptance by many professionals in the testing

field that test scores are not readily altered by the introduction of incentives.

General Predictions

The present research was designed to investigate in a controlled fashion the effects of incentives upon test performance. The investigation was cross-cultural in part and operated on the assumption that indigenous Hawaiians and Caucasians, as groups, have widely variant attitudes and reinforcement histories in regard to achievement.

Thus, while it was expected that both ethnic groups would perform better under incentive than under standard conditions, it was also predicted that introduction of incentives would have a greater influence upon the test performance of Hawaiians than the Caucasians. In the present study there was a hopeless, but unavoidable confounding of ethnic background, school, and socioeconomic status; most Hawaiians live in a Hawaiian community and are mostly poor, and similar enclaves of poor Caucasians are unavailable in Hawaii. On the basis of anthropological and some experimental evidence, there was also reason to believe that among Hawaiians, group incentives would be more potent than individual incentives while the reverse might be true for Caucasians; therefore, conditions of both individual and group incentives were incorporated into the experimental design.

Implications of Positive Findings

If it could be shown that incentives can appreciably influence test scores, and that different classes of incentives are differentially effective among children of different ethnic backgrounds, then there would be several implications for practical action as well as for the re-evaluation of traditional testing and diagnostic practices. Important decisions are made on the basis of test scores and spuriously low scores (due to "low motivation) could lead to stigmatic diagnostic labeling or inappropriate class placement. Program evaluations based on pre-posttest differences may be invalid in many cases due to "low motivation" during the testing phases of the program; conversely, "low motivation" during pretesting and "high motivation" during post-testing--after strong "rapport" had been established--could make a program look spuriously good. On the positive side, unequivocally significant results could pave the way for the development of refined assessment procedures which incorporate incentive conditions (for example, splitting tests into equivalent forms and administering one form under standard conditions, the second under incentive conditions).

At an individual level, certain students whose test scores increase markedly from standard conditions to conditions of incentive may be ill-placed in low ability or learning disability classrooms. Optimal learning climates

for such students may be in "regular" classrooms where incentives are provided. At a normative or cultural level, if certain response/incentive contingencies are related to improvement in test performance for large groups of students, there may be broad spectrum implications for educational planning. For instance, if it could be shown that indigenous Hawaiians perform "optimally" under testing conditions of group incentives, there might be good reason to introduce such concepts as team grades or group incentives into the network of Hawaiian education.

Use of Money as an Incentive

Money was used as the incentive in this study. Why? Was there any reason to believe that the promise and/or delivery of money would motivate all the subjects and to a similar degree? Practically speaking, it would have probably been impossible to get all subjects "optimally motivated," i.e., motivation held constant across subjects.

For several reasons, however, the offer of money was probably the general incentive of choice. First, money is a naturally occurring "token reinforcer," which can be redeemed or exchanged by an infinite variety of self-selected reinforcers. Surely, other incentives might have been more reinforcing in individual cases, but two facts precluded their use in this study. First, it would require a great deal of impractical effort not only to identify, but also to make available, idiosyncratic

reinforcers. Second, utilization of idiosyncratic reinforcers would still be no guarantee that motivation would be held constant or be optimal across subjects. The group use of money in the present study, nevertheless, was an attempt to get at a general-idiosyncratic incentive of one specific ethnic group--Hawaiians.

While the use of incentives may not cancel out all differences in motivation, their use may provide a means of reducing the variance in motivation between subjects. And it should probably be pointed out again that in the establishment of norms for most tests, no attempts have been made to measure the effects of motivation over and above the effects of ability or aptitude.

CHAPTER V

EXPERIMENT ONE: A PILOT STUDY

Due to two major confoundings, Experiment One was considered a pilot study and its results interpreted with reservations. First, there was a large difference between the Hawaiian and Caucasian groups in terms of latency between pre- and posttest. Second, entire classes of students were nested within the various conditions of the experiments, that is, entire classes--rather than individual students--were randomly assigned to each condition. The study is included here primarily because what was learned substantially influenced the subsequent research and also appears to have practical relevance.

Introduction

Many schools administer standard achievement tests to their students each year for purposes of class placement and determination of individual student's progress and general achievement levels. Experiment One attempted to show that such scores would go up if students were promised[1] money for good performance. The experiment

---

[1]The experimental manipulations of Experiment One were designed to determine whether the verbal promise of reward (money) in exchange for correct answers would result in improved test performance. In a technical sense, then, the experiment was designed to exert discriminative rather than reinforcement control over test taking behavior. That

also attempted to show that for Hawaiians promise of group reward would have more effect than promise of individual reward; also, that effects for Caucasians would be opposite, i.e., promise of individual reward would have more effect than promise of group reward.

Method

Subjects. An original Hawaiian sample of 220 subjects included the students in all of the seven sixth grade classes at a public elementary school in a low income, rural area on the Island of Oahu. All but a few students were of part-Hawaiian ancestry, and many resided on lands which are leased by the Department of Hawaiian Homelands to individuals of at least 50 percent Hawaiian ancestry. The final Hawaiian sample of 177 excluded students who did not take the posttest.

A Caucasian sample included 93 fifth and sixth grade students from three classes at a public elementary school

---

is, money was promised or offered for correct answers, but money was not actually dispersed contingent upon correct responses during the test period. Although the discriminative and reinforcing function of a stimulus are likely correlated highly, the distinction may be an important one (Staats, 1968); the promise of money for certain behavior may not always control behavior in the same manner as the dispensing of money following that behavior. For example, the promise of money for a reason and in a situation in which money had never been dispensed before might lack "credibility" and as a result not exert strong control over "striving" behavior; on the other hand, if money were actually administered in the situation, "credibility" could not be an issue.

in a lower-middle income area in Salt Lake City, Utah. Four other students who did not take the posttest were excluded.

Initially, the author had hoped to recruit additional subject populations from two private schools on Oahu, comprised of students of high-achieving Hawaiian and high-achieving, high-income Caucasian backgrounds. Unfortunately, permission to conduct research in these two schools was not granted.

Test Used. All subjects received both alternate forms, 4A and 4B, of the Scholastic Test of Educational Progress (STEP), a standardized 50-item multiple-choice, math achievement test published by the Educational Testing Service.

Experimental Design. Entire classrooms were randomly assigned to either a control (no-incentive), individual-incentive, or group-incentive condition. (One Hawaiian classroom was assigned to a free-time incentive condition, but this condition was excluded from the primary, between-school analysis.) Thus, the assignment of subjects was in a two factor, 3X2 design, with two levels of school or ethnic background (Hawaiian/Caucasian) and three levels of type-of-administration (no-incentive/individual-incentive/group-incentive).

All subjects received a baseline-pretest administered in a standard manner. All subjects were then retested

under either the control or one of the experimental conditions. The dependent variable in all cases was the difference between a subject's pretest score and his posttest score.

General Procedures. There were differences in general procedure for the Hawaiian and for the mainland Caucasian samples. In Hawaii STEP is administered yearly in September in public schools. Therefore, the scores from the regular Fall administration of STEP (Form 4A) were obtained from school records and used as the baseline-pretest measures for the Hawaiian students. Seven months later, in April, the students received an alternate form retest (Form 4B) under one of the several conditions. As in the pretest, STEP was group-administered to students by their regular classroom teachers in their regular classrooms--under conditions that approximated as closely as possible the testing atmosphere at the beginning of the school year.

The Salt Lake City public school system does not utilize STEP. Therefore, the Caucasian students received STEP math on two consecutive days--first Form 4A and then Form 4B--the first administration given under standard testing conditions per the STEP directions for proper administration. Thus, the period between pre- and posttesting was much shorter for the mainland Caucasian sample.

A second difference between treatment of the two subject populations was that the same teacher administered

STEP in all three conditions to the Caucasian subjects. This teacher was the regular daily math teacher for two of the classes and the regular daily art teacher for the third.

Instructions. All subjects were informed that the tests were being administered for purposes of the school, for determination of individuals' math achievement levels. The student-subjects were not told that they were participating in an experiment.

At the start of the baseline-pretest session, the classroom teacher read the test instructions from the STEP directions manual, and the test was administered in the standard manner.

On retest, subjects in the control group received the alternate form of STEP math in exactly the same standard manner as in the baseline-pretest.

On retest, subjects in the individual-incentive condition received the following additional instructions immediately following the standard test instructions:

> The test you are to take now involves something I want you to understand completely. Each of you will be able to earn money for doing this test. By working hard and reading each question, you will earn money. Each student will get 5¢ for every question he answers correctly. So, the more right answers you get, the more money you will make. For example, if (insert name of a student) gets 10 right answers, he will get 10 nickels to keep for himself. If he gets 15 right answers, he will get 15 nickels. You will be paid as soon as the tests are scored.

On retest, subjects in the group-incentive condition received the following additional instructions immediately following the standard test instructions:

> The test you are to take now involves something I want you to understand completely. Listen carefully. The class will be able to earn money for doing this test. By working hard and reading each item, you will earn money for the group to spend. Every student will get 5¢ for the group for each answer he gets right. All the money that everybody earns will be collected together into one large amount. Then the class can decide on how to spend the money on some group activity. You may want to buy snacks, cokes, have a class party, go to a movie together, or have a picnic. How you spend the money, as a group, partly depends on how much you earn. But everyone will share in the same amount. I will add up all the right answers that everyone gets, and the class will receive a nickel for every right answer. You will be paid as soon as the tests are scored.

On retest, subjects in the free-time condition, which included one class of Hawaiian students and no Caucasians, received the following additional instructions:

> The test you are to take now involves something I want you to understand completely. Listen carefully. For doing this test, each student will be able to earn free time during the school day for himself and his classmates. By working hard, by reading each item, and by answering questions correctly, you will earn time to do anything you want to do during school time. Every student will earn 10 seconds of free time for the whole class for each right answer he gets. All the free time will be added up into one large amount. And the whole class will use all the minutes to do whatever they want to do during the school day. For example, if (insert name of student) earns 3 minutes of free time for his right answers, everybody in the class will get 3 extra minutes to do whatevery they want to do. And if (insert name of second student) earns 2 minutes for his correct answers, that will be added to the 3 minutes and make five minutes of free time for everybody. In this way, all the minutes that each student earns will be added up,

> and everybody will share in the same amount of free time. You will get your free time as soon as the tests are scored.

Following completion of the experiment, the participating students were told that they had taken part in an experiment and each, including control subjects, was given one dollar.

Results

Table 1 presents the means and standard deviations of the difference scores for each of the school type-of-administration conditions. Excluding the free-time condition, the results of a two-way analysis of variance for unequal cell frequencies (Winer, 1962, pp. 241-244)--summarized in Table 2--fail to reveal any statistically significant effects due to type-of-administration. Subjects in the individual and group-incentive conditions did not improve in performance from pretest to posttest, relative to the subjects in the control condition. In addition, there were no school X type-of-administration interactions. The overall pre-posttest difference-scores were significantly greater for the Hawaiian subjects than the mainland group ($F=5.73$, $df=1/236$ $p < .01$), but this difference was likely due to the fact that the pre-posttest delay was six months for the Hawaiians and only a matter of days for the Utah subjects; the six-month interval provided the opportunity for considerable learning to have taken place.

Table 1

Means and Standard Deviations of Step Math Difference-Scores
Type of Administration

School (Ethnic Group)	Control (No Incentive)			Individual Incentive			Group Incentive			Free-Time Incentive		
	N	M	SD	N	M	SD	N	M	SD	N	M	SD
Hawaiians	65	1.98	5.48	37	2.38	4.92	47	2.83	6.13	28	0.89	4.24
Caucasians	31	1.42	3.86	33	0.18	4.98	29	0.52	5.65	-	-	-

Table 2

3X2 Analysis of Variance of Step Math Difference-Scores

Source	df	MS	F
School-Ethnic Group (A)	1	161.71	5.73*
Type-of-Administration (B)	2	4.89	0.17
A X B	2	16.92	0.60
Error	236	28.24	

*$p < .01$

Collapsing all incentive conditions--including the free-time subjects--into a single incentive condition, a second analysis of variance was conducted and is summarized in Table 3. Once again, the results indicate neither a type-of-administration effect nor any interaction effects. The use of incentive instructions did not result in test score increases in Utah or Hawaii.

Discussion

The results of Experiment One did not provide any support for the hypothesis that promise of reward can enhance test performance. While one might conclude that introducing incentives into the test situation is inconsequential, the failure to achieve significance could be attributable to (1) insufficient potency of the experimental manipulations, (2) an overly difficult examination, and/or (3) the fact that correct "guesses" may have contributed substantially to the error variance.

Instructional (Discriminative Stimulus) Control Versus Reinforcement Control. The promise of money in exchange for correct responses may not have exerted strong control over student striving behavior. In the absence of a history of being paid for correct responses in the classroom, the promise of incentive instructions (discriminative stimulus control) may have lacked "credibility" and as a result not controlled striving behavior to the extent that

might have been possible had money been actually dispensed immediately following correct responses.

Table 3

2X2 Analysis of Variance of Step Math Difference-Scores (Incentive Conditions Collapsed)

Source	df	MS	F
School-Ethnic Group (A)	1	80.60	2.970*
Type-of-Administration (B)	1	10.32	0.38
A X B	1	23.00	0.85
Error	266	27.14	

* $p < .10$

Multiple-Choice Versus Free-Response Tests. The test used in Experiment One was comprised of 50 multiple choice questions with four response options per question. As a result, an inevitable source of chance error was introduced into the analysis. That is, if each question were simply answered in a random manner, in the long run the mean expectation would be 12+ correct answers (25 percent) for each subject. A student's score would be expected to vary somewhat from test to retest simply on the basis of his "guesses" (each of which has as an *a priori* probability of 25 percent of being correct). For example, a student could have obtained two more correct answers--due to increased "effort"--under incentive conditions than

under standard-baseline conditions. This performance improvement could have been "washed out" or obscured by three "lucky guesses" during the standard administration which were "unlucky guesses" during the incentive administration. Of course, the aforementioned source of error is a nonsystematic chance one and should cancel itself out in the long run; nonetheless it adds necessarily to the error variance. This source of error is not present in the kind of test on which the test-taker must construct his own solutions (free or open-end responses) rather than select from several preselected alternatives.

Task Difficulty. On the school-administered Fall pretest, 51 percent (91/177) of the Hawaiian students obtained 30 percent or fewer correct answers--scores with a high probability of occurrence on the basis of chance alone. Indeed, the mean score for the entire Hawaiian sample--at 17.0--was only about five points higher than one would expect--in the long run--on the basis of chance alone (12.5 or 25 percent). This fact suggests the distressing possibility that on the basis of skill alone, the mean score for the Hawaiian students would have been approximately five correct responses out of a possible fifty.[2] To dramatize the point of this paragraph, it

---

[2]This observation is even more disconcerting when one realizes that these students had received the same or an equivalent STEP math test two years earlier in the 4th grade. STEP math comes in levels appropriate for use with 4th, 5th and 6th graders (Forms 4A, 4B) as well as

would be folly to offer a person even a million dollars to solve a calculus problem had he never advanced beyond basic math. The 50 word-problem questions on the STEP math exam fall within a very narrow range of difficulty and also require the test-taker to have relatively sophisticated reading skills[3]--a notable achievement deficit among Hawaiians.

While task difficulty could conceivably have had something to do with the fact that the Hawaiians showed no performance increments under incentive conditions, this explanation could not account for the nonsignificant pre-posttest changes among the Utah students who performed at significantly higher levels (pretest mean = 29.11) than the Hawaiian students (pretest mean = 17.02) ($t = 11.57$, $df = 268$, $p < .001$).

The Use of Standardized Multiple-Choice Achievement Tests in Hawaii. The results of the school-administered STEP math referred to above provide absolutely no

---

in levels appropriate for 6th, 7th and 8th graders (Form 3A, 3B). All fourth grade classes receive Form 4A or 4B. However, the test's makers recommend that "above average" sixth grade classes be administered Form 3A or 3B and "average" or "below average" sixth grade classes be given Form 4A or 4B.

[3]Sample STEP math item: "In New York they will visit the Empire State Building. Including the television tower, it is 1472 feet tall. To the nearest hundred feet, its height is (a) 1400, (b) 1470, (c) 1500, or (d) 2000?"

discrimination between the skills of the 51 percent of the Hawaiian students who received chance expectation scores although it would be folly to assume that all these students were functioning at the same level of math achievement. Thus for these children the results are meaningless other than to say that they all fall in the bottom half of math achievement among all sixth graders, and the bottom three percent of national norms.

At best, the results of standardized multiple-choice tests like STEP math--whose questions call upon complex reading skills and whose items fall within a very narrow range of difficulty--can be used for distinguishing gross levels of differential achievement. (Admittedly, finer meaningful discriminations are possible among groups of students whose scores are distributed broadly from very high to very low than are possible among groups of students whose scores are highly skewed.) However, such tests provide virtually no precise information as to what a child can or cannot do. And if a student does poorly simply because he didn't "try," there is no way to know or suspect on the basis of his score markings--unless, for example, he systematically marks the first choice to every question.

Use of free-response tests--on which the test-taker constructs his own solution rather than pick from pre-selected alternatives--seems far preferable to the use of

tests like STEP. First, free-response tests can provide considerably more information than multiple-choice tests about what a student can or cannot do. For example, on the basis of a child's free-response test answers, a teacher can often tell whether a student can do subtraction, multiplication, or "carry" in addition. Such information can serve as a precise basis for programming a student according to his current level of (math) functioning. All too often teachers "instruct" the entire class at the same level of difficulty when some students don't have the slightest idea what's going on and as a result are likely to be psychologically remote from the academic proceedings. If a child doesn't know how to "carry" in addition, how could he be expected to understand complex multiplication?

Second, if one uses free-response exams--such as the Sullivan Math Placement Test employed in Experiments Two and Three--it is often possible to determine errors resulting from "carelessness" or "low motivation." For example, one student--during baseline testing in Experiment Two--made errors on three problems in one-column basic addition but went on to get every single problem correct on the multiplication page. It goes without saying that this child has mastered basic addition and obviously did not try to do his "best" on the basic addition problems.

One might conclude from all this that the primary advantages of standardized multiple-choice tests over

free-response tests is that such scores can be compared to national norms and that they can be machine scored. In Hawaii, all public schools spend almost one entire week each year administering a series of multiple-choice aptitude and achievement tests (e.g., math, reading, social studies, listening, etc.), each of which takes about ninety minutes to administer. From the student's point of view there are no extrinsic consequences to this testing ordeal. In fact, most students never learn their test scores. What a punishing way to start the school year. Operating on the assumption that "work" in the absence of reinforcement is aversive, one might conclude that this week of testing is an excellent way to "turn off" students with respect to academics and schoolwork early in the school year--precisely at the time they should be getting "turned on."

CHAPTER VI

EXPERIMENT TWO

Experiment Two was addressed to the same general questions as Experiment One; namely, whether incentives would alter test scores and whether Caucasians and Hawaiians would be differentially influenced by individual and group incentives. In a general sense Experiment Two was a replication of Experiment One but, aside from a different subject population, Experiment Two was also distinguished by several important methodological differences. First, in order to gain greater control over the influence of extraneous variables, Experiment Two was conducted in a more laboratory-oriented atmosphere than Experiment One. Instead of being administered by regular classroom teachers in the natural context of the school day, Experiment Two was conducted in a specially prepared room by nonschool personnel who ostensibly were studying a math placement test for the University. Second, the test employed was not multiple-choice, but instead required the subjects to compute their own answers to each question. And while the subjects in Experiment One were merely promised money for correct answers, the subjects in Experiment Two also received money (five cents) immediately following each correct response <u>during</u> the test session.

Method

Subjects. Sixty-one Hawaiian students were recruited from a low-income area, rural Oahu elementary school--different from the school in Experiment One--and included all the Hawaiian ancestry students in the sixth grade, all of whom resided on Hawaiian Homelands. (About 90 percent of the students at this particular school are of part-Hawaiian ancestry.)

Seventy-five Caucasian subjects were recruited from the fifth and sixth grades of an elementary school in the second largest city on the Island of Oahu; the student population at this school come from a variety of ethnic backgrounds and socioeconomic levels.

Test Used. All subjects received the Sullivan Math Placement Exam, which was designed by its makers to provide teachers with a means of starting students at the appropriate book level in Sullivan Associates' Programmed Math. The test consists of 70 free-response math problems, including basic and advanced addition, subtraction, multiplication, division, fractions, and decimals. Each of the seven areas is represented by ten problems on a given page. The area heading for the page on which a student first makes more than two errors indicates the recommended book-level starting point in Programmed Math. Total scores are disregarded for purposes of placing a student in the Sullivan series.

Experimental Design. The design of Experiment Two was essentially the same as Experiment One--individual subjects being assigned in a 3X2 factorial design, one factor being school or ethnic background (Hawaiian/Caucasian), the second being type-of-administration (no-incentive control/individual-incentive/group-incentive). Individual subjects from each school were randomly assigned to either the control, individual-incentive or group-incentive condition. All subjects received the same "standard" administration as a pretest-baseline measure. All subjects were then retested under either the control or one of the two experimental conditions. In all major analyses the dependent variable was the difference between a subject's pretest score and his posttest score. Measures were obtained on raw scores, book-level scores and time spent taking the test.

General Procedures. Subjects were tested in groups--ranging in size from 12 to 15 persons. At each of the two participating schools, special rooms were prepared for the experiment, and the subjects sat in three rows of four or five. The same female graduate student read the test instructions to all students in each condition. A second adult female was present as proctor at all times, and a third adult female was present during all posttesting. All three experimenters were Caucasian.

The pretests were administered over a period of six days at each school, two testing sessions being conducted each morning. After all pretests were completed, the control subjects were retested first. On the following day the group-incentive subjects were retested and on the sixth and final testing day the individual-incentive subjects received their posttests.

During the posttest--in all conditions--each of the three adult experimenters stood behind one row of students. During the standard-control posttest, each served merely as a proctor; however, during the individual and group-incentive posttests, the experimenters took on an additional role.

During the posttest each of the three adult experimenters stood behind one row of students. Each had become familiarized with the correct answers to each problem and held an answer sheet in one hand. Each also carried an over-the-shoulder cloth bag containing several dollars in nickels. As each student completed a test item, he was instructed to raise his pencil in the air, at which point the experimenter for that row checked his answer; if correct, the experimenter either dropped a nickel into a jar in front of the student (individual-incentive) or into the "sharing jar" carried by the experimenter (group-incentive).

During the incentive condition posttests, students were required to do each page of problems in sequence, and once an answer was checked by an experimenter, it could not be changed. When one of the experimenters checked the last test problem or when a student signalled that he or she had finished (not all students answered all the problems), the experimenter marked down the time elapsed since the start of the test.

Instructions. The student subjects were not informed that they were going to participate in an experiment and were simply referred to the experimental testing room by their regular classroom teachers. At the beginning of the baseline-pretest session, the test booklets were passed out, and all students were read the following test instructions:

> May I have your attention please. We are from the University of Hawaii and are studying a math placement test. Your names have been selected by chance to take part in our study of this test. Please fill out the blank spaces on the front cover. Do not open your booklet until I tell you to.
>
> Attention, please. The test you are about to take is simply a math test. When I say "begin," open your booklet and start working on the problems. This is not a speed test. You will all have plenty of time to finish every problem. Work carefully and accurately. When you are finished, raise your hand and I will collect your test.
>
> If you finish early, please remain seated quietly until everyone else finishes. Please do not talk to anyone during the test.
>
> Any questions? Ready, open your booklets and begin.

Upon retest, the control subjects received the following similar instructions:

> May I have your attention, please. Fill out the information on the front cover. Do not open your booklet until I tell you.
>
> Today you are going to take Part 2 of the test you took last week. The test is a math test very similar to the one you took last time [actually the same test].
>
> When you are finished, raise your hand and I will collect your test. If you finish early, please remain seated quietly. Please do not talk to anyone during the test.
>
> Remember! This is not a speed test. You will all have plenty of time to finish every problem. Work carefully and accurately. Any questions?
>
> Open your booklets and begin.

Upon retest, the group-incentive subjects received the following instructions:

> May I have your attention, please. Fill out the information on the front cover. Do not open your booklet until I tell you. Today you are going to take Part 2 of the test you took last week. The test is a math test very similar to the one you took last time. However, the test involves a special procedure I want you to understand completely. Every person will get 5¢ for the group for each answer he gets right. When you finish working on a problem, raise your pencil so we can check your answer. If your answer is right, we will put a nickel into the sharing jar. If your answer is wrong, you will not earn a nickel for the group, and instead will receive a tap on the shoulder--which means for you to move on to the next problem. But do not move on to the next problem until each answer is checked by us.
>
> Each person is working for the group as a team. At the end of the test we will take the nickels out of the sharing jar and pass them out to the class. You will all share in the same amount. But the more right answers you get, the

more money you will earn for (yourself and) the group.

When you have finished with the test, raise your hand and we will collect your test booklet. If you finish early, remain seated quietly. Please do not talk during the test.

Remember! This is not a speed test. You will all have plenty of time to finish every problem. But once we have checked an answer, you can not go back and change your answer. Work carefully and accurately. (Any questions?)

Ready, begin.

Upon retest, subjects in the individual-incentive condition received the following instructions:

May I have your attention, please. Fill out the information on the front cover. Do not open your booklet until I tell you to. Today you are going to take Part 2 of the test you took a few days ago. The test is a math test very similar to the one you took last time. However, the test involves a special procedure I want you to understand completely. Each of you will receive 5¢ for each correct answer, and each person will be working only for himself. When you finish working on a problem, raise your pencil so we can check your answer. If your answer is right you will receive a nickel in the cup next to you on the table. If your answer is wrong, you will not earn a nickel for yourself and instead will receive a tap on the shoulder--which means for you to go on to the next problem. But do not move on to the next problem until each answer is checked by us.

Each person is working for himself and each person will get to keep all the nickels he has earned. The more right answers you get, the more money you will earn for yourself. When you are finished with the test, raise your hand and we will collect your test booklet. If you finish early, remain seated quietly. Please do not talk to anyone during the test. Remember! This is not a speed test. You will have plenty of time to finish every problem. But once we have checked an answer, you can not go back and change your answer. Work carefully and accurately. Any questions? Ready, begin.

At the end of the testing session in the group-incentive condition, the students were told that as soon as the earned money was divided up equally, each child would receive his equal share. At the end of the individual-incentive posttest, the money was placed in individual packets, and the students were told that they would receive their packet of money at the end of the school day. Thus, all students received their earnings to keep after completion of the entire experiment. Subjects in the control condition were each given a one-dollar gratuity.

Results. Figure 1 presents the mean raw difference-scores obtained by the Hawaiian and Caucasian subjects in each of the control and incentive conditions. The results of a two-way analysis of variance for this data are summarized in Table 4. The overall analysis of variance for the type-of-administration effect is significant at the .03 level ($F = 3.70$, $df = 2/130$), and indicates that the raw scores of subjects in the incentive conditions went up significantly from pretest to incentive posttest relative to the raw score increases of control subjects. The results of subsequent Newman -Keuls tests indicate that both the individual and group-incentive subjects had significantly higher mean raw difference-scores than the control group. There were no significant differences between the individual and group-incentive subjects however.

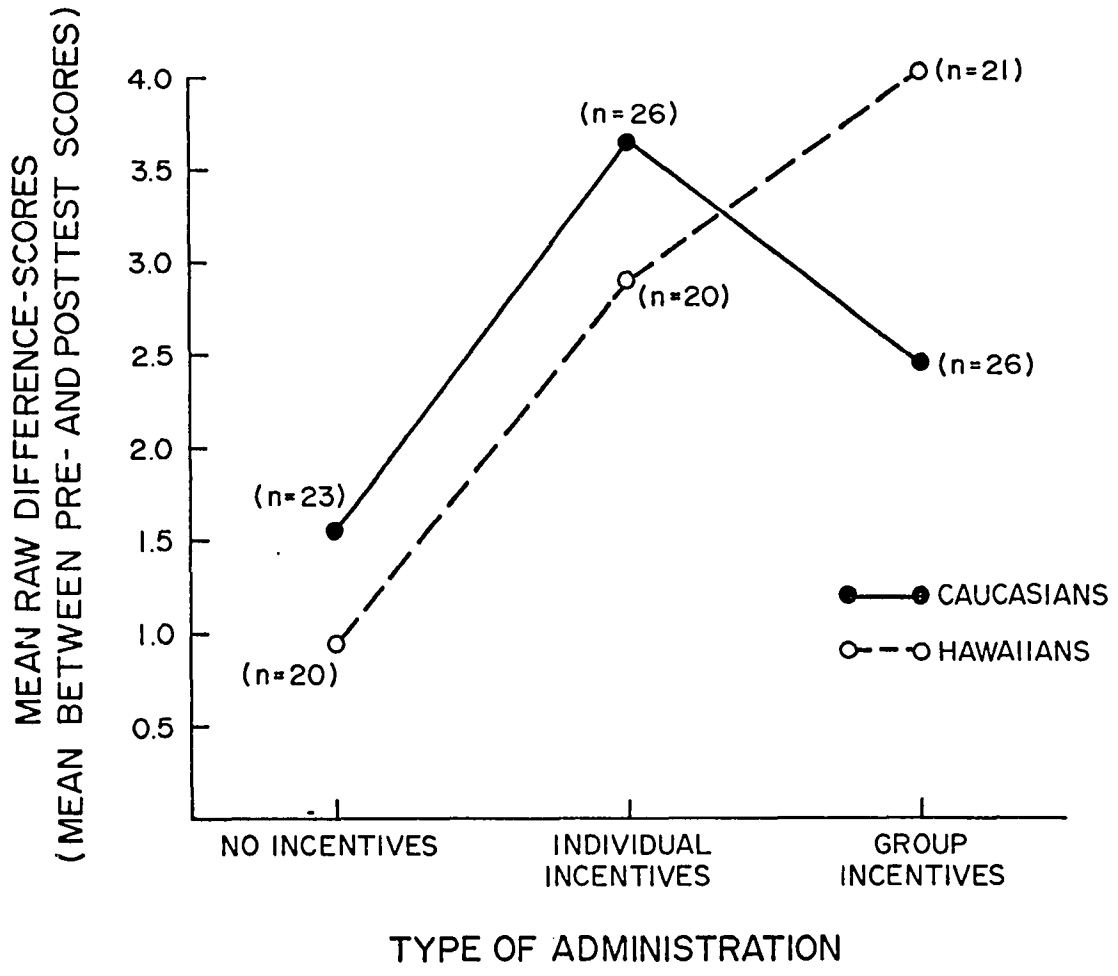

Figure 1. Mean Raw Difference-Scores Obtained by Hawaiians and Caucasians on Sullivan Math Placement Test Under No-Incentive (Control), Individual-Incentive, and Group-Incentive Conditions.

Table 4

Analysis of Variance of Sullivan Math Raw Difference-Scores

Source	df	MS	F
School-Ethnic Group (A)	1	0.437	0.03
Type-of-Administration (B)	2	59.77	3.70 *
A X B	2	17.23	1.07
Error	130	16.14	

*p .03

Although, as predicted, the mean raw difference-scores were greater for the Hawaiians in the group-incentive condition (M = 3.95) than in the individual incentive condition (M = 2.46), and greater for the Caucasians in the individual-incentive condition (M = 3.65) relative to the group condition (M = 2.46), statistically the interaction effect did not even begin to approach significance (F = 1.07, df = 2/130). Furthermore, although the use of incentives did result in performance increments, the Hawaiian and Caucasian groups did not differ significantly from one another in terms of absolute magnitude of test score increment (F = 0.03, df = 1,130).

Figure 2 presents the mean, book difference-scores obtained by the Hawaiian and Caucasian subjects in each of the control and incentive conditions. As indicated earlier,

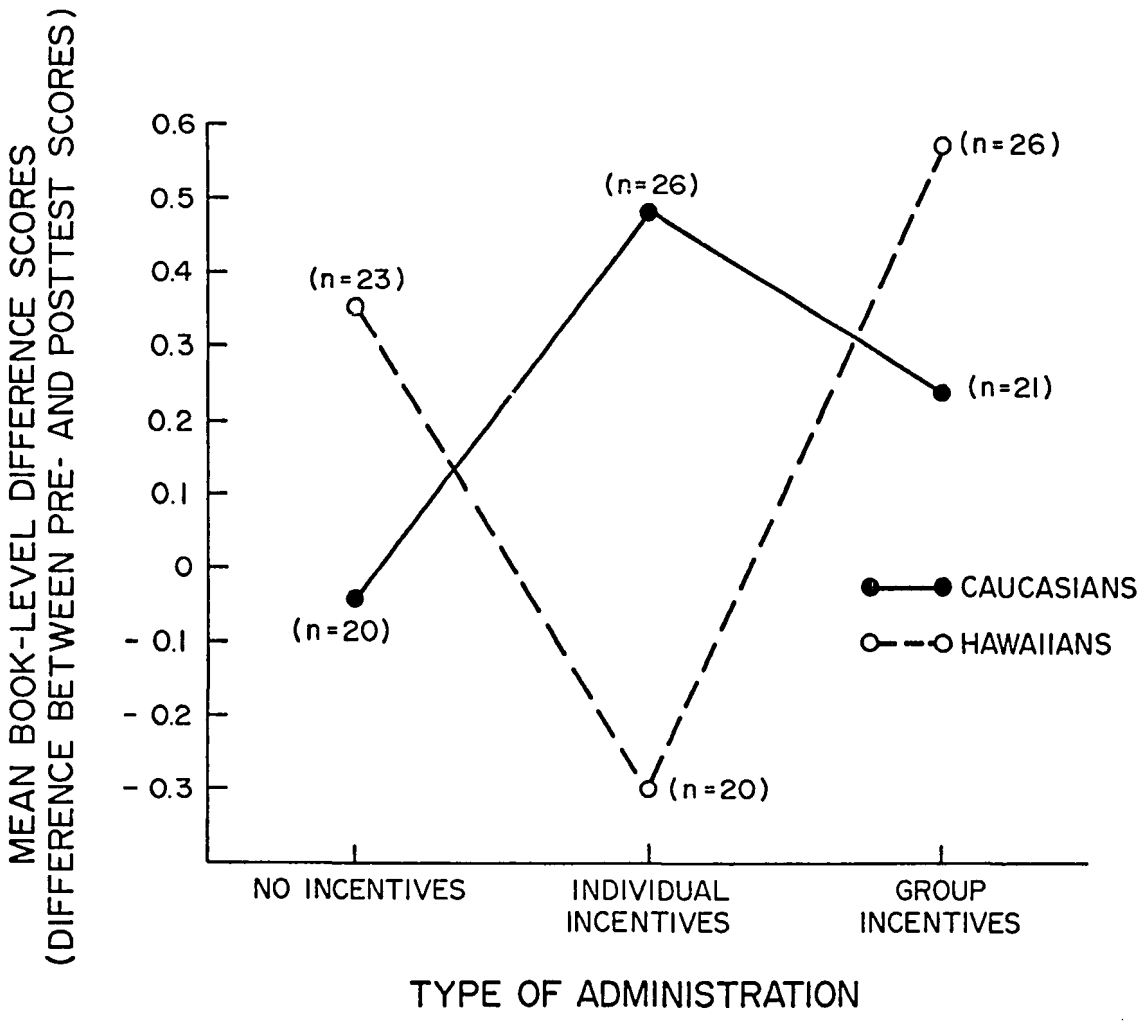

Figure 2. Mean Book-Level Difference-Scores Obtained by Hawaiians and Caucasians on Sullivan Math Placement Test Under No-Incentive (Control), Individual-Incentive, and Group-Incentive Conditions.

a student's book-level score represents his general level of math achievement and serves as an index for placement in the Sullivan Associates' Programmed Math. Book-level score is a cruder measure than total correct raw score since book-level scores may range only from one to seven, whereas raw scores may range from zero to 70. The results of a two-way analysis of variance for this data is presented in Table 5. Neither the main effect due to school or ethnic group nor the main effect due to type-of-administration reaches statistical significance. However, the school X type-of-administration interaction effect does reach marginal significance ($F = 2.63$, $df = 134$, $p < .10$). This marginal interaction effect seems to be primarily due to the superiority of the Caucasians over the Hawaiians in the individual-incentive condition.

Table 5

Analysis of Variance of Sullivan Math Book-Level Difference Scores

Source	df	MS	F
School-Ethnic Group (A)	1	0.01	0.06
Type-of-Administration (B)	2	1.21	0.65
A X B	2	4.88	2.63 *
Error	130	1.86	

* $p < .10$

Figure 3 presents the mean time difference-scores obtained by the Hawaiians and Caucasians in each of the control and incentive conditions. The results of a two-way analysis of variance for these data are presented in Table 6. The overall analysis of variance for the type-of-administration effect is highly significant ($F = 29.77$, $df = 2/130$, $p < .001$). Control subjects actually spent a mean 10 minutes less time working on the posttest than the pretest ($M = -10.15$) whereas individual incentive subjects ($M = 12.35$) and group incentive subjects ($M = 15.31$) spent considerably more time working on the incentive posttest than the pretest.

Table 6

Analysis of Variance of Sullivan Math Time Difference-Scores

Source	df	MS	F
School-Ethnic Group (A)	1	5,014	0.258
Type-of-Administration (BP)	2	579,650	29,768*
A X B	2	7,041	0.362
Within	130	19,472	

* $p < .001$

Results of Newman-Keul tests show that the subjects in the individual and group-incentive conditions both obtained significantly higher time difference-scores than subjects

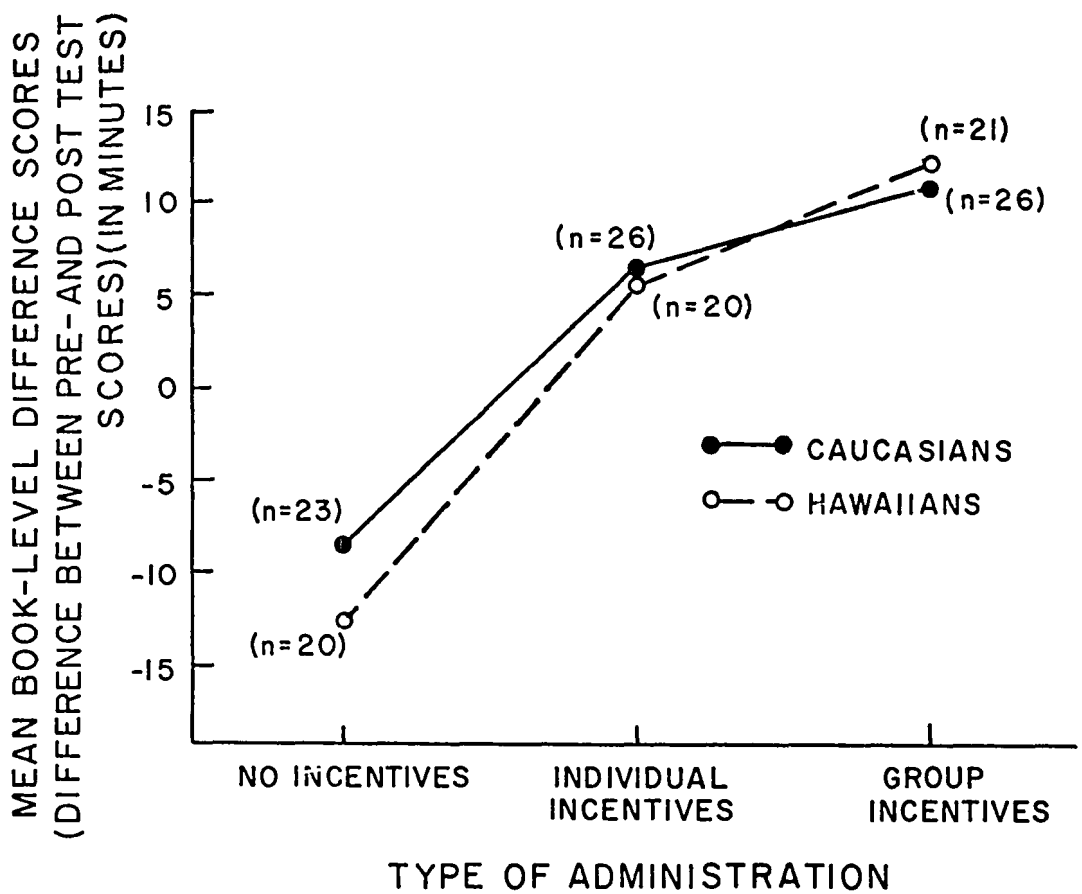

Figure 3. Mean Time Difference-Scores Obtained by Hawaiians and Caucasians on Sullivan Math Placement Test Under No-Incentive (Control), Individual-Incentive, and Group-Incentive Conditions.

in the control conditions--the differences in both cases being significant well beyond the .01 level. Although the Hawaiian and Caucasian subjects both had greater time difference-scores in the group than in the individual incentive condition, the post hoc difference between these two groups was not significant.

It seemed plausible that the potential magnitude of a subject's difference-scores was determined to some extent by his baseline level of responding. For example, if a student's total true knowledge of math did not go beyond basic addition, he would be unable to solve subtraction or multiplication problems--no matter what the incentive. And, in this regard, it may be important to note that the Hawaiian subjects (pretest $M = 34.80$) obtained far fewer correct responses than the Caucasians (pretest $M = 47.97$) ($t = 6.73$, $df = 134$, $p < .001$). Therefore, it was decided to consider analyses of covariance on the data so as to statistically hold constant pretest scores (Winer, 1962, pp. 605-606). First, however, correlation coefficients were computed in order to specify the extent of the relationship between pretest scores and difference-scores. Contrary to expectation, the pretest raw scores do not correlate significantly with the raw difference-scores ($r = -.16$, $df = 130$); therefore, the plan to conduct an analysis of covariance on the raw score data was abandoned. On the other hand, pretest time scores do correlate highly significantly--in the

negative direction--with time difference-scores ($r = -.34$, $df = 130$, $p < .001$). Shorter pretest time scores were associated with greater increments in time spent from pretest to posttest.

Figure 4 presents the weighted means of the time difference-scores for the Hawaiians and Caucasians in the control and incentive conditions. The summary of the analysis of covariance on this data is presented in Table 7. It can be easily seen (compare with Figure 3 and Table 6) that holding pretest time score constant has the effect of greatly magnifying the obtained differences.

Table 7

Analysis of Covariance of Sullivan Math Time Difference-Scores

Source	df	MS	F
School-Ethnic Group (A)	1	43,594	3.19 *
Type-of-Administration (B)	2	713,833	52.16 ***
A X B	2	63,515	4.64 **
Error	130	13,684	

*$p < .05$

**$p < .02$

***$p < .001$

First, as with the analysis of variance the overall analysis of covariance indicates that the main effect due

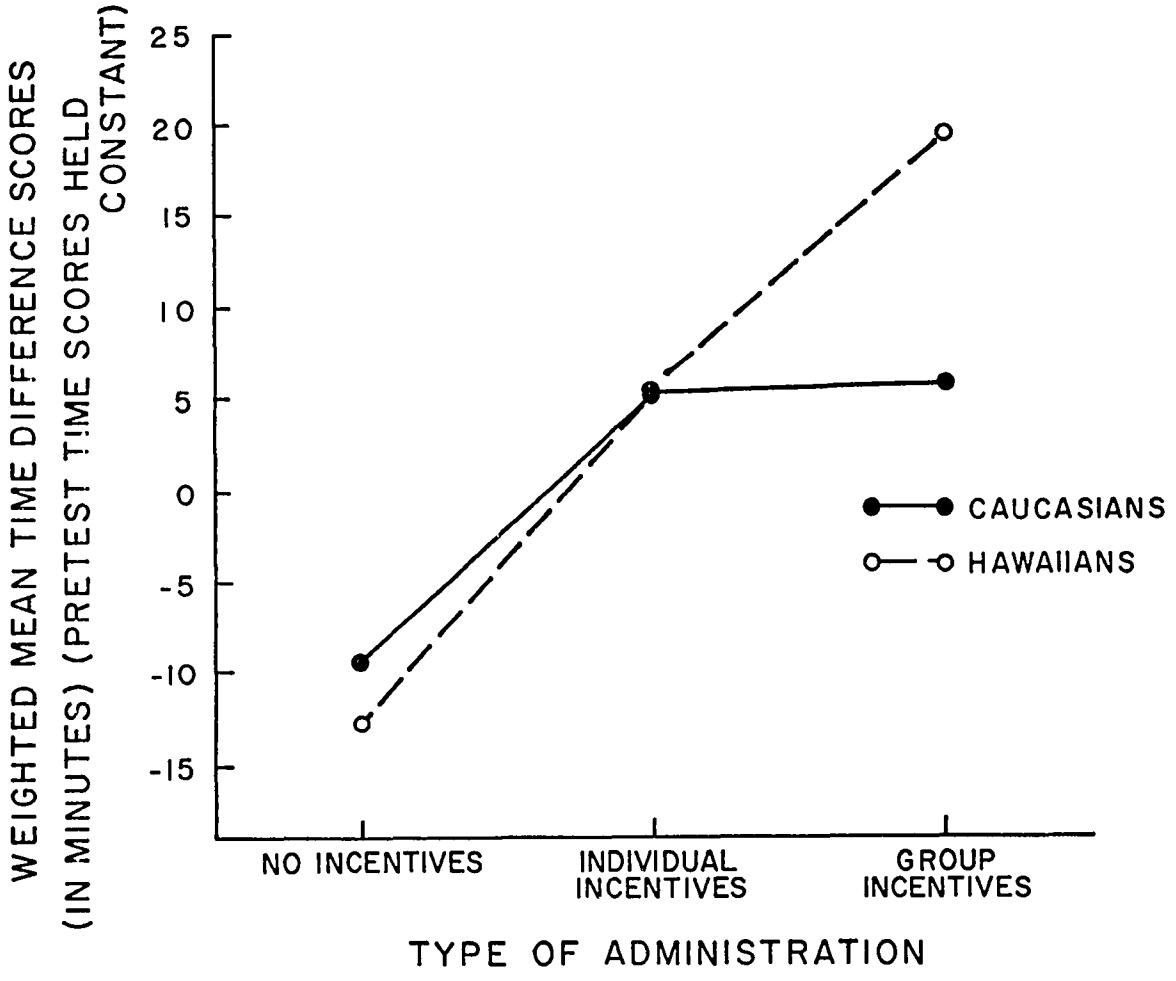

Figure 4. Weighted Mean Time Difference-Scores Obtained by Hawaiians and Caucasians on Sullivan Math Placement Test Under No-Incentive (Control), Individual-Incentive and Group-Incentive Conditions (Pretest Scores Held Constant).

to type-of-administration is highly significant ($F = 52.16$, $df = 2/129$, $p < .001$). And post hoc comparisons again show that the time difference-scores in both the incentive conditions are significantly larger than the difference-scores in the control condition ($p < .001$). But, in addition, with pretest time scores held constant, the post hoc difference between the individual-incentive condition (weighted $M = 4.53$) and the group-incentive condition (weighted $M = 13.53$) also reaches statistical significance ($p < .01$). Inspection of Figure 4 shows that this difference is attributable primarily to the large time difference-score of the Hawaiians in the group-incentive condition (weighted $M = 20.34$).

The analysis of covariance main effect due to ethnic group also attains statistical significance ($F = 4.64$, $df = 2/129$, $p < .02$). With pretest time scores held constant, the Hawaiian subjects spent more time working on their posttests (relative to their pretests) than did the Caucasian subjects.

The analysis of covariance on the time difference-scores also reveals a significant ethnic group or school X type-of-administration interaction effect ($F = 3.19$, $df = 1/129$, $p < .01$). Inspection of Figure 4 shows that this interaction is due primarily to the larger time difference-scores of the Hawaiians (relative to the Caucasians) in the group-incentive condition and to a small extent due to the fact

that in the control condition the Hawaiians had greater post-test time decrements than the Caucasians.

Discussion. Both Hawaiian and Caucasian groups spent more time on their math exams and obtained higher scores when they were given money contingent upon correct answers than under relatively standard testing conditions. The experimental manipulations produced greater achievement (higher test scores) as well as greater perseverance (longer time spent working). That the students were "trying harder" under the incentive conditions was also evidenced anecdotally by their greater "involvement" during the incentive test sessions. Although no reliable behavioral measures were taken, the children seemed quieter, less distractible, and were rarely seen squirming in their seats. Furthermore, that the students were highly concerned about "success" or "getting a good score" was evidenced by the frequent smiles, sighs of relief, and laughter when the drop of a nickel signified a correct response and the frequent moans and grimaces when a tap on the shoulder signified a wrong response.

One should take caution, however, not to overgeneralize from these results and consider first certain parameters of the experiment. The results refer to scores obtained in a relatively artificial experimental environment, to a test which is graduated in difficulty, to a testing situation in

which no time limit is imposed, and to incentive conditions under which the contingencies are specified to the students and incentives are actually delivered during the test session.

The results of the analysis of covariance--with time spent on pretest held constant--suggest that the subjects worked longer during the group-incentive posttest--relative to pretest--than during the individual-incentive and control posttests. Standing alone, these results provide slight support for the notion that "peer pressure," if channeled, can be a powerful servant of academic motivation. The greater "perseverance" observed under group-incentive conditions was due largely to the performance of the Hawaiian subjects, and again, provides mild support for the hypothesis that Hawaiians will "try harder" when working toward group goals than individualistic ones. The results of the analysis of covariance also suggest that Hawaiians spent more time working on their posttests--relative to pretest--than did the Caucasian subjects. This finding supports the original hypothesis that Hawaiians would be more influenced by incentives than would the Caucasians, but standing alone, once again provides only suggestive support of the hypothesis.

With the exception of the analysis of covariance on the time difference-scores, described above, the predicted ethnic group X incentive condition interactions were generally

not borne out in a statistical sense. For example, while the Hawaiians did obtain higher mean difference-scores (raw, book-level, and time) in the group-incentive condition than in the individual-incentive condition, no resounding statistical significance was achieved. Retrospective analysis of the experimental manipulations, however, suggests that Experiment Two may not have been a particularly powerful test of the relative efficacy of individual versus group incentives. In the individual-incentive condition, dispensing of coins for individualized performance was designed to provide not only an individualized goal, but also to arouse a "sense of competition" which, if aroused, was expected to interfere somewhat with the performance of the Hawaiian students. There is some reason to believe, in retrospect, that a potent experimental manipulation--concerning individual versus group-incentives--was not accomplished. Overlooked was the fact that although subjects in the individual-incentive condition did in fact receive money (token reinforcement) for individual accomplishment, there was no control over the "back-up" reinforcers for which the token earnings could be redeemed. For example, although a student did earn his tokens for "individual" performance, there was nothing to stop him from spending the fruits of his labors on his friends as well as himself. It would have been very interesting to know how the Hawaiian and Caucasian subjects spent their

earnings, to what extent they spent it on themselves and to what extent they shared their profits with their friends.

In a replication of this experiment, a "purer" distinction between individual and group-incentives might be accomplished by assuring that the "back-up" rewards are used solely in an individual or group fashion and/or increase the salience of the individualistic-competitive versus shared-cooperative distinction. For example, in the individual condition, the experimenter might advise the subjects that those individuals with the highest scores would have their names posted on the bulletin board and receive their earnings individually in front of an audience. Or, the choice of "back-up" rewards might be specified in advance. For example, in an individual condition, the students with the ten best scores might earn individual time off from school or unpleasant classwork; in the group condition, every member on the "team" with the highest total team score might earn the privilege together.

CHAPTER VII

EXPERIMENT THREE

The Youth Correctional Facility on the Island of Oahu has an educational program for its delinquent inmates, which is conducted by State of Hawaii certified teachers. The Sullivan Math Placement Exam is administered at the beginning of the school year or when adolescents are first committed and again at the end of the school year for purposes of measuring a youth's progress and for general program evaluation. The nature of the situation thus permitted a school-administered, natural environment, partial replication of Experiment Two. The school counselor agreed to readminister the Sullivan Exam for experimental purposes.

Method

Subjects. The original sample included 25 adolescent inmates, ages 13-18, all but one of whom were male. Fifteen and ten youths were assigned randomly to the experimental and control groups respectively. The final sample of 21 excluded two experimental and two control subjects who were lost when they were released from the Facility just prior to experimentation.

Experimental Design. The pretest, posttest, difference-score design was used again in Experiment Three. The difference-scores of the experimental and control subjects

were compared by independent groups, t test analysis.

Several subjects who received incentives for correct responses were tested later, again under relatively standard conditions. A repeated measures analysis was used for comparing the scores obtained by these subjects during the different testing phases.

General Procedures. At the end of the school year, in June, the school counselor group-administered Sullivan Math in a standard manner and without time-limit to all of the youths in the educational program at the Facility. Two months later, with no school program having intervened, the counselor retested the youths, in pairs--first the control subjects and then the incentive subjects. The youths were told that they were being retested because the Facility needed additional information about their math achievement levels.

After receiving the August posttest, under standard conditions, seven control subjects were retested again--within a few days--under incentive conditions. In addition, scores from a return-to-baseline measure were obtained from 14 subjects when they were tested two months later--under relatively standard conditions--as part of the regular Fall testing program at the Correctional Facility.

Instructions. All subjects were familiar with the test, having taken it at least once previously. The control subjects were simply told to work the problems, and

that it was not a speed test--they would have ample time to finish every problem.

The subjects receiving money-incentives were read the following instructions:

> This test involves a special procedure I want you to understand completely. You will receive five cents for each correct answer. When you finish working on a problem, raise your pencil so I can check your answer. If your answer is right, you will receive a nickel next to you on the table. If your answer is wrong, you will not get a nickel and instead will receive a tap on the shoulder--which means for you to go on to the next problem. But do not move on to the next problem until I check your answer.
>
> You will get to keep all the money you have earned. And the more right answers you get, the more money you will earn.
>
> Remember, this is not a speed test. You will have enough time to finish each problem. But once I have checked an answer, you cannot go back and change your answer.

Results

When retested with response contingent rewards, the experimental subjects solved a mean 5.07 more problems per subject than they had solved during the baseline-pretest sessions--a 11.7 percent raw score improvement. On the mean average, the control subjects solved 0.75 more problems per subject during retest--a 1.4 percent increase. A t test comparing the change scores of the experimental and control subjects shows that the raw scores of the experimental subjects went up from pretest to posttest to a

significantly greater extent than the raw scores of the control subjects ($t = 1.92$, $df = 19$, $p < .05$).

The experimental and control group's book scores went up a mean 0.92 and 0.62 book levels, respectively. Results of a t test comparing these means were not significant ($t = 0.44$, $df = 19$).

During the testing sessions in August the counselor recorded the length of time each student spent taking the test, for which no time limit had been imposed. The subjects receiving money for correct responses worked a mean 57.00 minutes before handing in their test. The control subjects worked a mean 41.60 minutes. Analysis by an independent t test indicates that these means are significantly different ($t = 2.61$, $df = 19$, $p < .01$).

Seven control subjects were retested within a few days, this time with incentives added, and their time-elapsed scores were recorded again. This time they spent a mean 53.29 minutes taking the test, compared to a mean 38.86 minutes just a few days earlier. Results of a matched t test indicate that the time increment is highly significant ($\underline{t} = 3.23$, $df = 6$, $p < .01$). Thus, it is clear that these subjects voluntarily worked on their tests for longer periods of time when correct-response-contingent rewards were made available.

Figure 5 shows graphically the mean scores obtained during the baseline-pretest; incentive-posttest, and return-to-baseline follow-up for the 14 subjects who were tested in

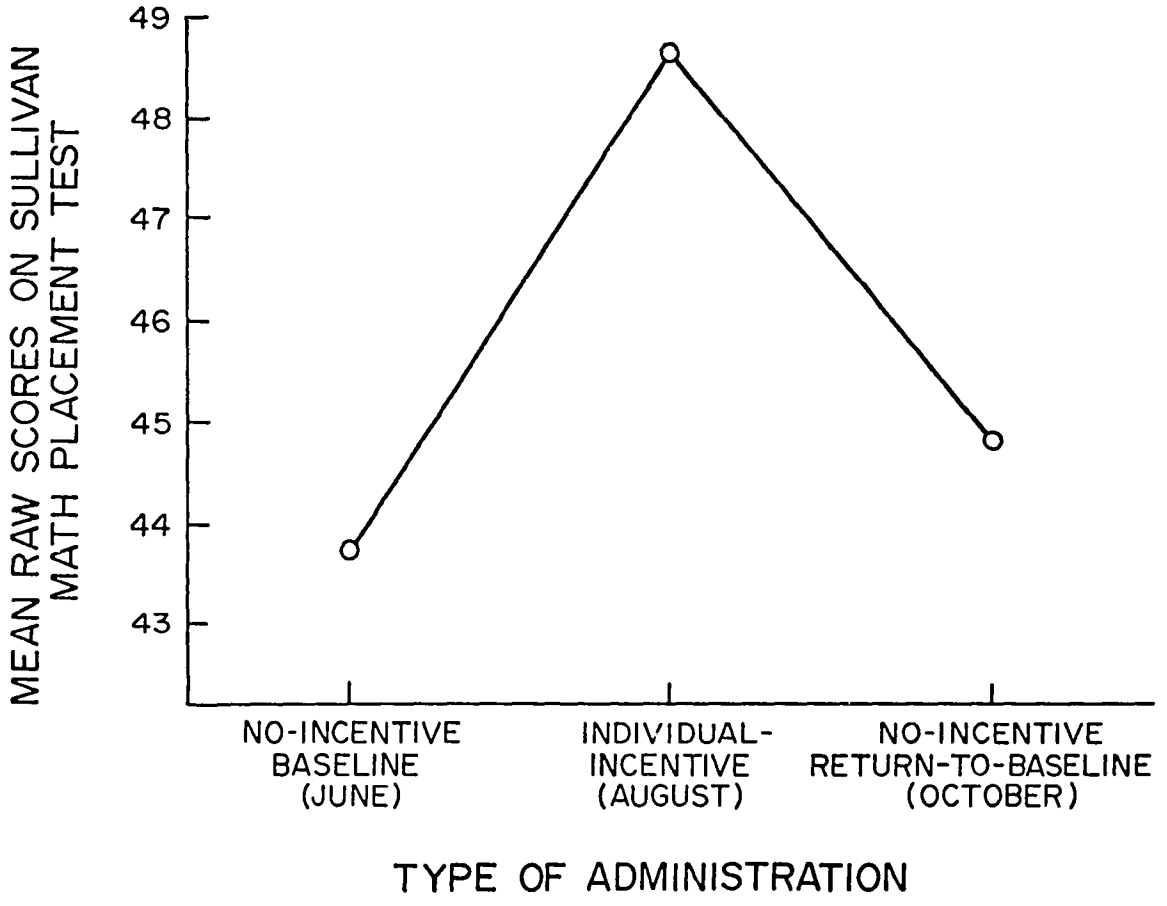

Figure 5. Mean Raw Scores Obtained on Sullivan Math Placement Test by 14 Subjects Given Repeated Measures Under No-Incentive Baseline Conditions, Individual-Incentive Conditions, and No-Incentive Return-to-Baseline Conditions.

all three phases. The mean raw scores of these 14 subjects went from 43.64 to 48.64 from the baseline to the incentive session--and then dropped to 44.93 during the return-to-baseline session. The overall results of a single factor, repeated measures analysis of variance on this data--summarized in Table 8--are significant at the .025 level ($F = 4.61$, $df = 2.26$). The results of subsequent Newman-Keuls tests reveal that these 14 subjects obtained significantly higher scores during the incentive test session than during both the baseline test session and the return-to-baseline test session. The scores obtained during the baseline and return-to-baseline sessions were not significantly different however.

Table 8

Analysis of Variance of Sullivan Math Raw Scores
(Repeated Measures)

Source	SS	df	MS	F
Between Subjects	6076.75	13		
Within Subjects	721.37	28		
Type-of-Administration	188.75	2	94.38	4.61*
Residual	532.62	26	20.49	
Total	6798.12			

*$p < .025$

Ten of the 14 subjects received lower scores on the return-to-baseline testing phase than on the previous incentive phase, two subjects obtained higher scores, and two showed no change. The 10 subjects who "relapsed" had a mean eight fewer correct solutions on the return-to-baseline test than on their previous incentive test!

Discussion

The results of Experiment Three, consistent with those obtained in Experiment Two, provide strong support for the contention that some students do not try to do their best under standard testing conditions, and that provision of appropriate rewards-delivered contingent upon correct responses will elicit improved test performance, both in terms of score and "perseverance." This conclusion is strengthened by the results of the repeated measures analysis which shows that after the money-incentives were withdrawn, scores dropped significantly.

Interestingly, the subjects at the Correctional Facility made greater test-score gains--relative to pretest--than their counterpart subjects from the two elementary schools in Experiment Two. The 20 subjects in Experiment Three who were tested under incentive conditions obtained a mean 5.00 more correct responses when receiving money for correct answers than under standard, pretest conditions, compared to a mean, raw score increment of 3.21 for the 90 experimental subjects (individual and group incentive) in

Experiment Two. While these scores may not be technically comparable, a t test analysis shows that their difference reaches marginal significance ($t = 1.79$, $df = 108$, $p < .10$, two-tailed). The apparently greater improvement of the Correctional Facility subjects should probably not be considered surprising since a great many in the delinquent population have been "turned off" to school related matters for some time, and it is likely that those youths find little or no reinforcing value in achievement stimuli.

The results of the present study, obtained in a naturalistic setting, may have direct and immediate implications for educators and others concerned with academic programs which involve youths with long histories of academic failure, school avoidance, or under-achievement--whether they are considered "juvenile delinquent," "emotionally disturbed," or "mentally retarded." As a specific example, the results suggest that test scores obtained in youth correctional facilities, in general, may in many cases not be valid. What a disconcerting possibility! Program evaluators in such settings might seriously consider some alterations in their methods of academic assessment--such as those suggested below.

It would be a relatively simple matter to split a test like Sullivan Math into equivalent halves or to construct an equivalent alternate form. Upon admission to a special program, a child might receive one form under standard

conditions and the other under incentive conditions. Such procedures might yield a plethora of useful information. First, the examiner could be reasonably confident that he had elicited an accurate measure of the student's current level of functioning--with respect to the dimension measured. Second, discrepancies in scores and social behaviors (e.g., perseverance or distractibility) between the standard and incentive conditions might provide the examiner with information about the reinforcement value of achievement stimuli and academic work for the student. And considering the student's behavior in the test situation as a sample of the kind of behavior likely to be emitted in a wide range of academic situations, the diagnostician would be in a position to predict what training conditions under which a child might function more productively or more poorly. For example, a large performance discrepancy in favor of the incentive condition might indicate the need for extrinsic motivational support--say, in the form of a token economy or simple provision of tangible rewards for academic accomplishment. Later, retesting under standard and incentive conditions might provide information over and above an accurate estimate of a child's progress in the program. If, for example, a student exhibited a large discrepancy in score and task-oriented behavior under the two pretest conditions, but on the posttest he performed equally well (in terms of score and diligence) under both the standard and incentive

conditions, then, the examiner would have some basis for concluding that achievement stimuli had taken on some self-rewarding value for the student; this might be a basis for recommending phase-out into a regular classroom atmosphere. On the other hand, a student who continued to "fool around" and score poorly under standard posttest conditions might profit from a continuation of extrinsic props for academic work--even if his "incentive" posttest score showed that considerable academic progress had been made in the program.

"Perseverance" Under Intrinsic and Extrinsic Test Conditions as a Measure of Motivation. The more highly significant differences obtained with the time dependent variable than with the score dependent variable in Experiments Two and Three suggest that a "perseverance" measure--such as time spent working voluntarily--may be more sensitive to intrinsic-extrinsic motivational differences than an achievement score measure. The length of time that a student will voluntarily remain in a test (work) situation--with and without rewards available--might, in itself be a fairly good predicter as to how well he will adjust to the "intrinsic" atmosphere of most regular classrooms. While "striving" or "persevering" is by no means immediately synonomous with "achieving," "perseverance" not only translates eventually into achievement but also brings on social accolades from the pleased teacher.

Feedback--Reinforcement Confounding. In Experiments Two and Three there was no way to separate the effects of reinforcement from the possible effects of informational feedback. At the same time that subjects were being paid for right answers and left unpaid for wrong answers, they were simultaneously being given feedback about the quality of their performance. It is conceivable that this information could have served a guiding and self-correcting function and as a result have been responsible for the improved performance under the incentive conditions. That is, one might argue that feedback as to accuracy, alone, may have been sufficiently responsible for the obtained results--that the delivery of monetary reinforcement was redundant and unnecessary. Of course, even if a feedback-only condition had been included, the picture might not be much clearer since one could still argue that feedback itself serves a reinforcing function--at least for some students. While unequivocal answers to these speculations are presently unavailable, it seems unlikely that informational feedback could have been responsible for the highly significant increase in elapsed time under the incentive conditions--unless of course feedback was in fact serving a motivating-reinforcing function; in which case the extrinsic incentive contention would continue to receive support.

CHAPTER VIII

GENERAL DISCUSSION

Recommendations for Future Research

The results of the present investigation are sufficiently promising to warrant further study of the effects of incentives on test scores. Such research is currently under way by the author and is designed to overcome certain limitations of the present research. Some procedural outlines of the in-progress study are described below and represent a recommended research design model. The study will include subjects with a history of academic difficulties; specifically, the subjects will be students referred to a learning disability clinic for diagnostic evaluation. Each subject will be tested under both standard and incentive conditions during a single, individually-administered testing session. Each of several subtests comprising a learning disability test battery will be split into equivalent, alternate-form halves. One test battery of shortened subtests will be administered under incentive conditions with pennies dispensed on a correct response-contingent basis.

Half the subjects will receive the incentive administration first, while the rest of the subjects will receive the standard administration first. Half the subjects will be promised pennies for correct responses and

paid while taking one of the shortened test batteries; the remaining subjects will be paid for their performance on the other shortened test battery. Incentive effects will be determined by comparing a subject's incentive condition score with his standard condition score on each of the subtests.

A feedback only condition will also be included in order to sort out the effects of monetary incentives over and above the possible effects of informational feedback. Subjects in the feedback only condition will be informed--in a neutral manner--each time they make a correct response, but they will receive no money.

Differential Diagnosis and Prescription in Education

The research model described immediately above may have practical extensions in many educational settings, as already implied in the Discussion Section of Experiment Three. Tests administered under both "intrinsic" and "extrinsic" conditions may occasionally yield results of significant diagnostic-prescriptive value to the psychological examiner at the school. It is suggested here that the term "motivational disability"--an equivalent of "low intrinsic achievement motivation"--be applied to students who score significantly higher under incentive testing conditions than under standard test conditions. Many such students may be inappropriately diagnosed as being retarded or as having

learning disabilities--on the basis of test scores obtained under "standard" conditions.

Adding the diagnostic classification "motivational disability" (perhaps a more innocuous sounding term would be better) and developing a measure of it could have profound implications for the diagnoses and treatment indications of children referred for testing because of academic difficulties. Inclusion of this new classification would open the way for a four-fold classification, useful in distinguishing students with learning problems, from those with mild retardation from those who simply seldom find anything rewarding about academic work. This four-fold classification system is presented in Table 9 and is described below.

Learning Disability--No Motivational Disability (LD-No MD). Students with true learning problems or disabilities--learned or organic--are typically distinguished on the basis of a jagged profile on a battery of tests measuring intellectual abilities, achievement, speech and

Table 9

Proposed Diagnostic System

		Learning Yes	Disability No
"Motivational Disability"	Yes		
	No		

hearing and perceptual-motor functions--scores on some subtests being average or high, but scores on others being normatively low. Students placed in this category would obtain essentially the same test scores and profile under both incentive and standard conditions. Figure 7 presents a hypothetical test profile of a student who would clearly fall within this category. Remedial treatment for such individuals might simply involve appropriate programming of materials in the area(s) of the child's difficulty and in many cases--given the availability of appropriate audio-visual materials--might not require removal of the child into a special class.

Learning Disability--Motivational Disability (LD-MD). Students placed in this category would also be distinguished on the basis of a jagged test profile obtained under both standard and incentive conditions. While the shape of the profile would remain jagged under both conditions, these students would obtain generally higher scores under incentive conditions. Figure 8 depicts a typical profile for a student wh would fit this classification. Such students are probably the ones who are most appropriately placed in special classrooms with small teacher-student ratios. These students need not only individualized work in sensory-perceptual remedial areas, but also require an environment with enriched opportunities for reinforcement (e.g., a token economy.

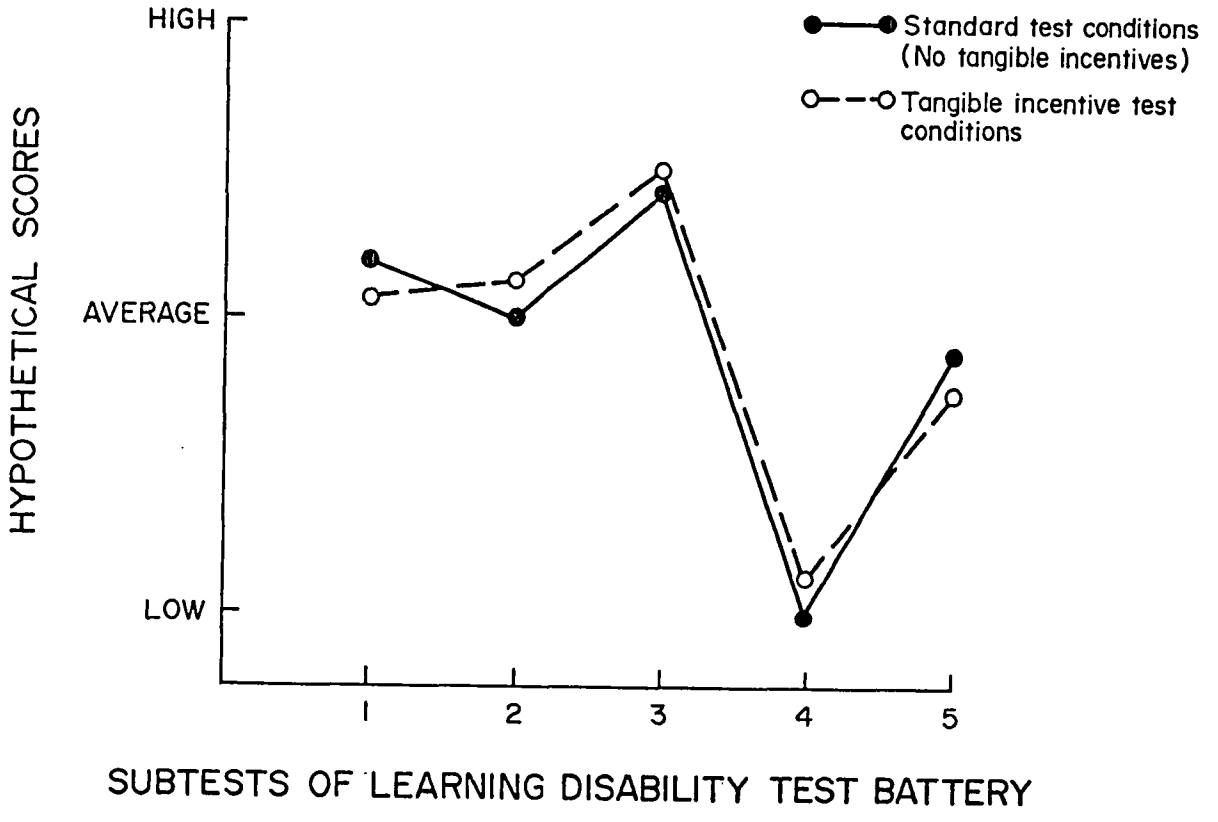

Figure 6. Hypothetical Test Profile Obtained Under Standard and Incentive Conditions of Child Diagnosed as Learning Disability--No "Motivational Disability".

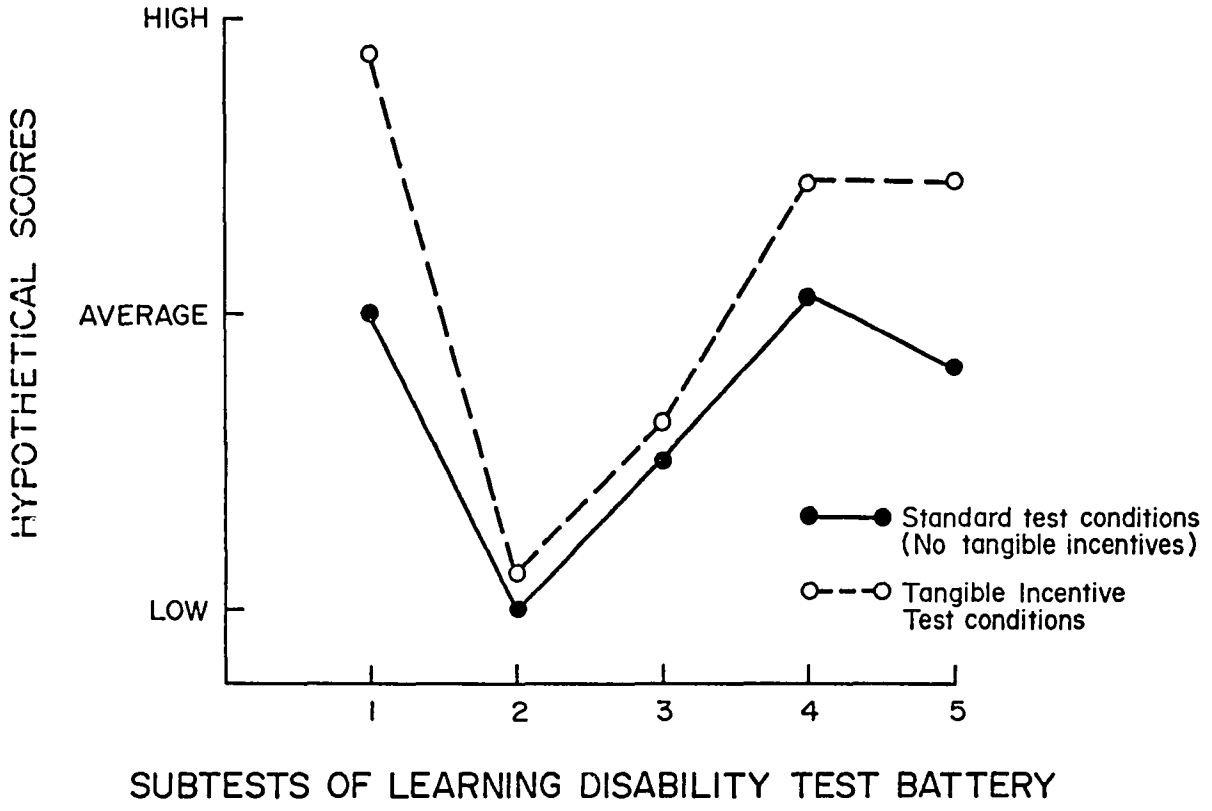

Figure 7. Hypothetical Test Profile Obtained Under Standard and Incentive Conditions of Child Diagnosed as Learning Disability--"Motivational Disability".

No Learning Disability--Motivational Disability (No LD-MD). Students who fall in this category would be the ones most likely mislabeled as being retarded or as having learning disabilities. A study by Kubany and Sloggett (1971) mentioned earlier describes such a case. Under standard testing conditions a generally low or jagged learning disability type profile would be obtained, but under incentive conditions scores would go up and the profile would flatten out--as indicated in the hypothetical test profile in Figure 9. Many of these students would probably profit from special motivational environments--for example, a token economy--but no sensory-perceptual remediation activities would be indicated; all that would be necessary in this regard would be to program academic materials in accordance with the student's current level of functioning.

Many students wh would fall into the "motivational disability" categories are probably diagnosed frequently as "hyperkinetic," "minimally brain damaged" and/or as having a "short attention span--due to central nervous system immaturity." Of course this kind of labeling creates a considerable amount of consternation among behaviorists since numerous studies have shown that children so labeled will "attend" and "persevere" for long periods of time, given appropriate conditions of reinforcement (Birnbrauer et al., 1965; Kubany, Bloch, and Sloggett, 1971; Patterson, 1965). Probably the worst consequence of such labeling is the

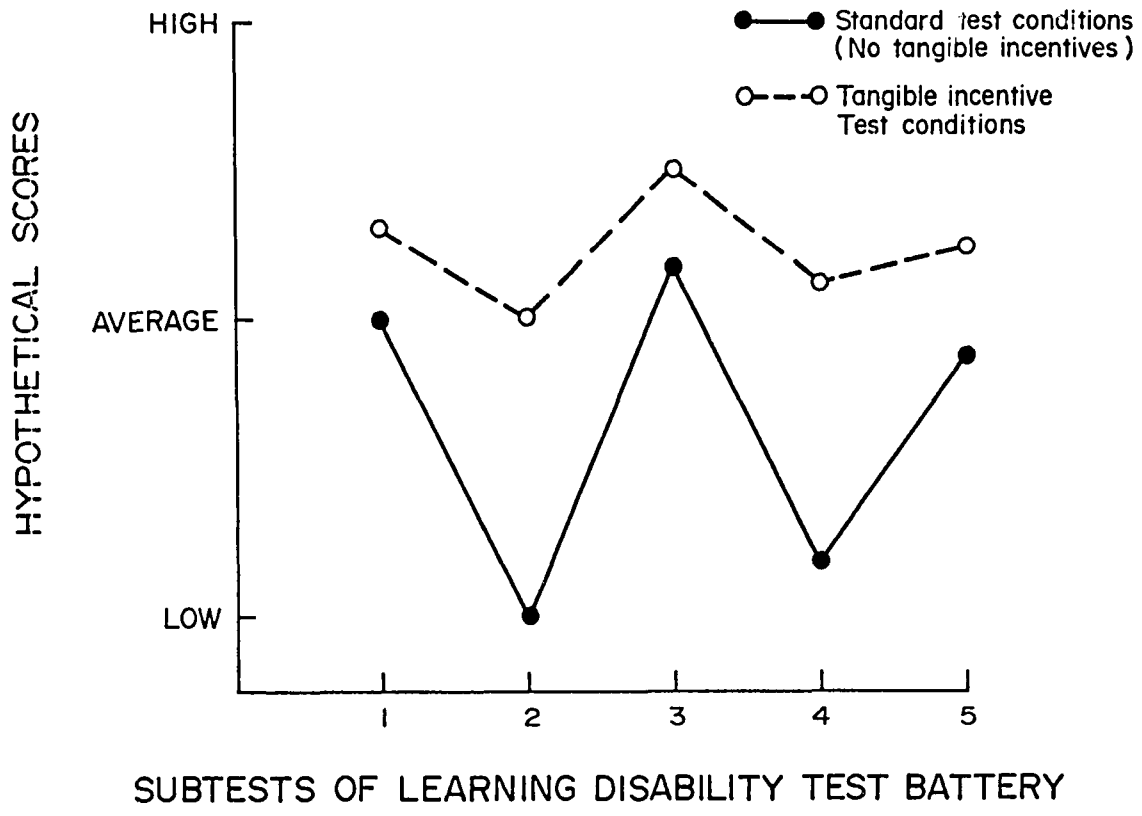

Figure 8. Hypothetical Test Profile Obtained Under Standard and Incentive Conditions of Child Diagnosed as No Learning Disability--"Motivational Disability".

"give-up-itis" which afflicts the trainer after he finds out that the child has been given a diagnostic necklace of organic involvement.

It should be reemphasized that the subjects in Experiments Two and Three spent considerably more time on their exams (less "hyperactivity," longer "attention span") under incentive than under standard conditions. It is the opinion of the author that far fewer students will be described as being constitutionally hyperactive or having short attention span--for neurological reasons--if they are tested under incentive as well as standard conditions.

No Learning Disability--No Motivational Disability (No LD-No MD). Students who fit this classification would receive similar relatively flat profiles under both standard and incentive conditions. Figure 10 shows two representative profiles, one of a child diagnosed as retarded, the other as average in every testing respect. No apparent treatment is immediately indicated for the "average" child other than to possibly look for social-situational crises in the child's life, conduct a functional analysis of the teacher-child social interactions within the classroom (Bijou, Peterson, and Ault, 1968) or possibly to have the teacher re-evaluate her reasons for referring the child for testing. For the retarded child, the two obvious choices of treatment would be individualized programming in the regular classroom or placement in a class of retarded children.

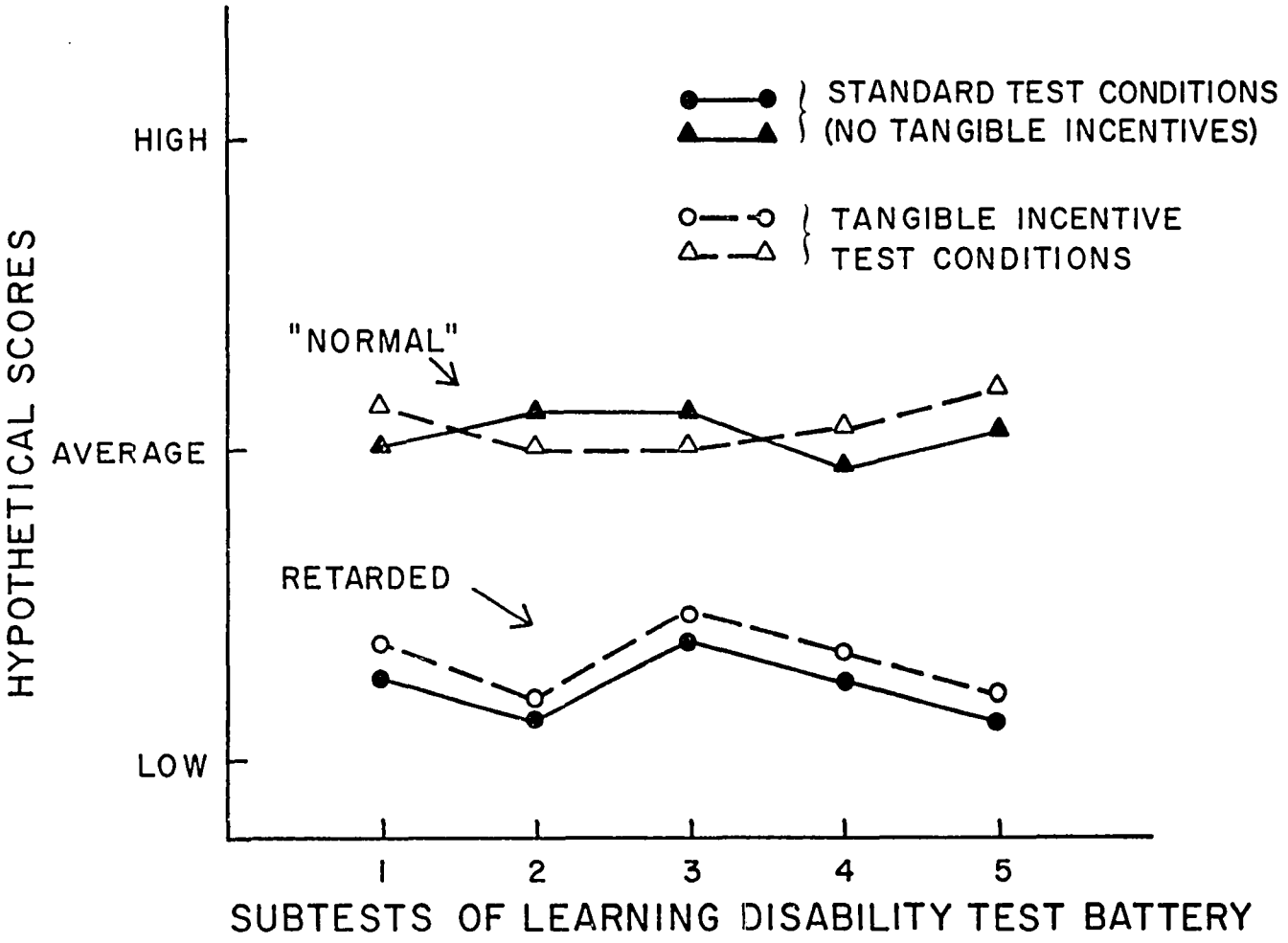

Figure 9. Hypothetical Test Profiles Obtained Under Standard and Incentive Conditions of Children Diagnosed as No Learning Disability--No "Motivational Disability".

It is recognized that the four-fold classification and remediation scheme can only apply in individual cases as an oversimplification, and that the use of stereotypic labels gives rise to expectations, not always positive, which are often fulfilled (Meichenbaum, Bowers, and Ross, 1969; Rosenthal and Jacobson, 1968). Nevertheless, while terms like "learning disability," "retardation," and "emotional disturbance" do often instill the trainer with a sense of futility or "limits" and imply that it's the child's fault when he doesn't function adequately, use of the term "motivational disability" is meant to imply that "there's nothing basically (organically) wrong with this kid. All we have to do is program him and channel his motivation." Here the trainer accepts the responsibility for the child's success or failure. Furthermore, despite its simplification, the four-fold classification scheme may represent a broad organizational framework for a complex child assessment model for clinical psychology. Based on learning principles, and crediting primarily the work of Arthur Staats (1970; 1971 a, b), the proposed model is described below.

Staats' A-R-D System

Staats (1970; 1971 a, b) has recently articulated a developmental-learning conception of motivation which could provide a theoretical rationale for the four-fold classification system described above. Staats calls his theory of human motivation the A-R-D system; the letters A, R and D

stand for the three functions that a stimulus may acquire--Attitudinal (emotional), Reinforcing, and Discriminative.

First, a stimulus may elicit a positive or negative Attitudinal (emotional) response. For example, the sudden appearance of a teacher or test (an achievement stimulus) before a student would likely elicit an emotional response in him--making him feel good or positive or possibly making him feel bad, anxious or negative. Second, the same teacher or test could also function as a Reinforcing stimulus. Teacher praise and/or a high test grade following "striving" (instrumental) test behavior would increase the likelihood of this behavior occurring again in the future. On the other hand, teacher criticism and a low test grade following "striving" test behavior would decrease the likelihood that the child would try to do well on a test in the future. Finally, the teacher or test could also serve a Discriminative or controlling function. The teacher or test might elicit or bring on strong approach responses from the student. For example, the student might come to school early to get ready for the test or for the opportunity to engage in social interaction with his teacher. On the other hand, if the teacher or test elicits a strong negative emotional response from the student, he might cut class on the day of the test or "not pay attention" to the teacher (a psychological avoidance response) when she lectures.

Thus, the same stimulus may serve three different motivational functions. Stimuli which elicit positive emotional responses in an individual (A function) will not only reinforce behaviors after they have occurred (R function) but will also "attract" him (D function). Conversely, stimuli which elicit negative emotional responses will not only negatively reinforce behavior by their removal but will also "repel" the person.

According to Staats, the formation of an individual's A-R-D system is dependent largely upon classical rather than instrumental conditioning. The system originates through pairing of previously neutral stimuli with stimuli that elicit emotional responses on an unlearned basis (e.g., food and water). When an attitudinal stimulus is paired with a new stimulus that is neutral in a rewarding-emotional sense, the new stimulus will acquire the three A-R-D functions. The system is seen as interactional; a change in the value of one function affects the other two functions. For example, changing the emotional response of a child toward his teacher from negative to positive would not only change the reinforcing value of the teacher for him, but the teacher would come to elicit different behaviors from him--for example, paying closer attention to her lectures. Elaboration of the system proceeds through extensive first order and higher-order conditioning with language subsequently playing a major role in new conditioning.

Staats (1971a, pp. 238-247) details various ways in which the learning situation, as a complex stimulus, can come to take on A-R-D values which foster or interfere with the development of pro-learning attitudes and self-reward systems, with respect to achievement. For example, when a child experiences "punishment" in a learning situation--due, perhaps, to setting of unrealistic standards by parents who communicate their disapproval when these standards are not attained--the punishing stimuli will elicit negative emotional responses in the child. And

> . . . each time an emotional response is elicited, it will be conditioned to all the stimuli present at that time. Thus, it must be expected that a child punished in a learning situation will learn a negative emotional response to the learning situations, the conditioned emotional response will become very general--it will be elicited by all learning situations that share common features with the other.
>
> This emotional conditioning will have general effects inimical to the child's learning. It should be remembered that the child will have learned to escape and avoid in various ways any stimulus that elicits a negative emotional response and thus has punishing functions. That is, any behavior that escapes the aversive stimulus will be learned through the reinforcement of the removal of punishment. Any behavior that avoids the stimulus will be learned in the same way. In general, then, when learning situations come to elicit negative emotional responses (anxiety) in the child, the child will learn behaviors that escape and avoid situations (Staats, 1971a, p. 239).

Staats (1970) addresses himself to what he calls "deficits" or "inappropriacies" in an individual's A-R-D system. "Abnormalities in conditioning history can provide individuals with an A-R-D system that has deficits such that

stimuli which should elicit positive or negative attitudes do not do so" (p. 136)--for example, teacher praise or a high grade not eliciting a positive emotional response from a student. Stimuli which should not elicit positive or negative attitudes but do so are seen as inappropriacies in the system. "The A-R-D system may be inappropriate such that stimuli which should not elicit an attitude response (either positive or negative) have come to elicit such a response" (p. 136--for example, the sensations produced from sniffing glue or stolen property as positive stimuli and a book or "teacher" as negative stimuli. In the cases of the glue and the stolen property, such stimuli may elicit strong "striving" responses which are incompatible with "striving" for achievement stimuli.

"Motivational Disabilities" as A-R-D System Defects. Students with "motivational disabilities" may be seen as having deficits or inappropriacies in their A-R-D systems. For these children, at least certain stimulus aspects of the test situation elicit negative emotional responses, are negative reinforcers, and control various psychological and behavioral escape and avoidance behaviors. The child is "forced" into the testing situation; if he had a choice, he would no doubt leave or would never have arrived in the first place. For some children with "motivational disabilities" the test stimuli, per se, may elicit strong negative emotional responses--test-learning situations

having been previously paired far more frequently with failure and punishment than with success and reward. With others, who may even have experienced a certain amount of "success" and achievement in the past, the negative emotional stimuli to be avoided may not be the test, per se, but rather the adult-teacher stimulus or the "staying seated" stimulus or the "hard work" stimulus. Or the competing attraction of highly positive peer stimuli may lead the child to "strive away" from the test situation. (These latter children frequently get labeled as "emotionally handicapped" and would fall into the MD-No LD category.)

In any event, for one reason or another "doing well" stimuli do not have sufficient discriminative-reward value to outweigh the aversiveness associated with the striving behaviors necessary to obtain them.

However, it is suggested that students with deficits or inappropriacies in their A-R-D system (motivational disability) <u>will</u> strive in test-learning situations in order to obtain <u>extrinsic</u> stimuli with strongly positive emotional-reward value. Such children can learn and perform in a variety of ways if the obtained rewards outweigh the aversiveness involved in obtaining them. Such children will "strive" if an <u>extrinsic</u> stimulus (e.g., money) is introduced and made available--contingent upon "test striving" responses.

Learning Disabilities as Instrumental Behavior Deficits. There is no universally accepted definition of learning disability (Bateman, 1966), and the term is usually defined so broadly as to be scientifically or operationally meaningless. For example, in the charter issue of the Journal of Learning Disabilities, Barsch (1968) refers to learning disability as "a term to be applied to any learner who fails to benefit from an existing curriculum into which he has been placed" (pp. 13-14). Usually the term refers in specific instances to a child's current level of achievement functioning on certain sensory, perceptual, and/or motor tests; to say one has a learning disability refers to a description of behavior rather than an explanation of behavior (Staats, 1963). Learning disability can be operationally defined in terms of a child's profile of scores on a learning disability test battery on which some scores meet or exceed the child's age norms while some scores fall below these norms. In one sense, then, learning disability could be viewed as "selective retardation." Admittedly, a learning disability test battery generally yield scores which tap a greater number of skill dimensions than obtained from general achievement tests (e.g., math or reading). The label presumably refers to what a child doesn't know which prevents him from acquiring new skills in the essentially nonindividualized atmosphere (e.g., the

same reading text for everyone) of many classrooms.[1] The label does not refer to how these behavioral deficits should be ameliorated. In sum, learning disability classification refers to a child's skill deficits and sometimes also to the specification of content level at which to begin remediation or training in instrumental behaviors. It is evident that the individual who has deficits in instrumental behaviors (LD or retarded, of course) may seldom obtain positive reinforcement in contiguity with achievement stimuli, thereby preventing the development of a positive A-R-D system with respect to achievement.

Interaction of Motivational Disability (A-R-D Defect) and Learning Disability (Instrumental Behavior Deficit). While an individual's A-R-D system will heavily determine the type of instrumental behaviors he will display, "there are additional interactions between the A-R-D system and the individual's instrumental behaviors, and moreover, the relationship between the two may go back and forth, each affecting the other and in turn being further affected by the changes produced in the other" (Staats, 1970, p. 139). However, changing a child's achievement stimulus A-R-D system from negative to positive would not be expected to result in academic progress where the instrumental behaviors

---

[1]If all students received individualized assessment and curriculum prescription, the terms "LD" and "MRE" (mild retardation educable) would lose much of their meaning.

necessary to obtain these stimuli are not in the child's repertoire. While altering an individual's achievement stimulus A-R-D system from negative to positive would change the child's general behavioral orientation toward achievement stimuli from avoidance to approach, the change would likely be shortlived unless the student has or is given the instrumental behaviors necessary to obtain these stimuli. Over a period of time the value of such stimuli would likely become negative again if attempts to obtain them are met with failure, frustration and disapproval from others.

Thus the plight of a child seen as having both MD and LD is particularly distressing because even if achievement stimuli come to elicit positive A-R-D values, the child would still find himself unable to obtain such stimuli--unless he were programmed for success--due to instrumental behavior deficits. In a very real sense, treatment of the LD-MD child would meet with difficulties similar to those confronted by Annon (1971) in his A-R-D analysis and therapeutic treatment of a homosexual man who had deficits in his A-R-D system with respect to adult heterosexual sex stimuli. After treatment had resulted in changing the valence of such stimuli (females) from neutral to positive, the man still found himself incapable of achieving such stimuli because of behavioral deficits in courting behavior. The essence of his dilemma is captured in the following quote

from Annon's patient: "I want to--and I know what to do when I get there--but how the hell do I get them to bed?" (p. 368) In a manner of speaking, this man came to therapy with both an MD and an LD, and elimination of the former was clearly only half the battle.

Application of Staats' A-R-D analysis to the MD-LD classification system gives rise to the speculation that students who fall into the LD-No MD category are primarily the ones diagnosed LD at an early age and then placed in special training environments emphasizing individualization and programmed success--thus giving the child the opportunity to develop pro-learning attitudes. Children diagnosed as LD in later years have probably experienced so much failure, frustration and criticism because of their behavioral deficits that it seems implausible that there would not also be associated inadequacies in their A-R-D system (MD).

While it is relatively easy to see how an LD (instrumental behavior deficit) can lead to an MD (A-R-D inadequacy) the reverse sequence may also occur--with MD contributing to LD. For example, a highly disruptive first grader (MD) with an age-adequate academic repertoire (no LD) could easily get caught in "vicious cycles of interactions . . . in which the defective A-R-D system leads to undesirable behaviors which lead to further defective developments in the A-R-D system, which again produce additional undesirable behaviors" (Staats, 1970, pp. 140-141). And, "after a few years of

school attendance when the conditions of learning are not appropriate for the child, he will not have acquired the behavioral repertoires acquired by more fortunate members of the class, whose previous experiences have established an adequate motivational system" (Staats, 1970, p. 140). With a history of "striving away" from learning situations--thus falling farther and farther behind--the child may indeed develop a learning disability--defined in terms of his test scores.

No "Motivational Disability"-No Learning Disability. What about the student who is assessed as having neither an MD nor LD? One should not automatically assume that the referring teacher completely erred in her appraisal of the child although occasionally a teacher may form an overgeneralized negative impression on the basis of one or two isolated aberrations on the child's part. An A-R-D analysis suggests a number of other possibilities. While the child's A-R-D system and current level of instrumental behaviors may be considered adequate, an analysis on the basis of A-R-D principles suggests that the child's "malfunctioning" is rooted in the specific learning situation in which he finds himself. For example, a particularly punitive teacher, as a stimulus, may acquire such negative A-R-D value that a child's poor classroom performance may be directly related to the child's psychological and/or behavioral attempts to escape or avoid the teacher. The child would not be

expected to pay attention (a psychological approach response) to a teacher whose verbalizations and very appearance elicit strong negative emotional responses. Another possibility is that the child will readily "strive" for positive teacher consequences but has not yet established an adequate self-reward system to maintain his work behavior. It should be noted that in an individual assessment session, adult social reinforcement ordinarily comes forth readily, and that a mild A-R-D achievement stimulus deficit might only become apparent if the examiner discontinued or withdrew social feedback from the student. In the case just described, the teacher may either have few social approval behaviors in her classroom repertoire or tend to dispense her social favors noncontingently. Obviously, in the case of the No MD-No LD child, a behavioral analysis of teacher-child interactions in the classroom is indicated. Another possibility is that the child has appropriate instrumental behaviors in his repertoire, but that the teacher has not clearly specified classroom rules or assignments so that the child will know unequivocally what responses on his part will obtain positive A-R-D stimuli. Still another possibility is that the training stimuli used by the teacher have extremely low discriminative-reward value or are simply too difficult for the child (can you imagine how boring and how difficult it would be to pay attention to a calculus lecture, especially if you had never had college algebra?).

The A-R-D Function of Language

Staats (1971a, b) has cogently spelled out how words or verbal stimuli acquire and possess A-R-D motivational functions. Spoken words "will elicit an emotional response and thus serve as either rewards or punishments, and also control 'striving for' or 'striving away from' behaviors" (1970a, p. 208).

> One of the powerful functions of language is the motivational function where new learning of behavioral skills can occur through the use of rewarding or aversive word stimuli . . . words delivered by other people mold our behavior according to the principle of reinforcement. Much of the child's school learning, for example, is maintained by verbal reward. The same is true for his learning of social behaviors. . . . (Staats, 1970a, p. 210)

However, it is important to understand that words originally are neutral stimuli and only take on their motivating functions after being paired with positive and negative rewards in the process of higher-order classical conditioning. Words as conditioned stimuli acquire their rewarding or aversive qualities from being paired with other stimuli which have such qualities. Recognizing this fact, one becomes aware of potentially large individual differences in the extent to which certain classes of verbal stimuli--for example, instructions, praise, and criticism--have discriminative-reward value. In fact, many students with "motivational disabilities" with respect to achievement stimuli may also have "motivational disabilities" or A-R-D defects with respect to certain verbal stimuli which play a

vital role in the development of academic and social skills. For example, individuals may vary greatly in the extent to which verbal praise elicits positive emotional responses, reinforces behavior and brings on or controls behavior aimed at obtaining praise. A child would have a marked A-R-D deficit if verbal praise did not elicit in him a positive emotional response, nor reward and control "striving" behaviors of various kinds. As another example, "constructive criticism" (e.g., "That's not good enough," "I've seen you do better") may function in diametrically opposite ways for different children. For one child such criticism may function discriminatively to contol his attending and task-oriented behavior--because in the past "increased striving behavior" has served to remove a criticism stimulus or bring on a praise stimulus. For another child the same criticism stimulus may serve not only to disrupt ongoing "striving" behavior but also to elicit a variety of "inappropriate" behaviors, such as tantrums, fidgetiness or reduced attention span--which this child has learned will serve to remove or terminate the criticism stimulus. As a last example, any teacher can tell you that her verbal instructions (e.g., "pay attention," "keep quiet") may exert total control over the behavior of some ("good") students--who have had a history of receiving reinforcement contingent upon obeyed instructions--but function as neutral stimuli (she might as

well be speaking a foreign language) to others, the "disruptive" ones, the ones with "motivational disabilities."

A Clinical Psychology Model for Child Assessment

Staats' A-R-D conception of intrinsic and extrinsic motivation which also recognizes the important motivational role of language may serve well as a conceptual scheme for ordering the academic-behavioral difficulties of children--on the basis of which can be made sophisticated decisions about the forms and sequence of treatment. However, assessment procedures which serve such a learning-principles conceptualization must of necessity depart from traditional testing practices which have better served psychodynamic ends.[2]

While some behaviorists have suggested that assessment take place in the natural environment where the child is experiencing difficulty (e.g., the home or school) and where he can be observed unobtrusively (Bijou et al., 1968), this kind of assessment has its limitations. For example, under such conditions it would be very difficult--especially within a short period of time--to arrange for manipulations in the

---

[2]In an article titled, "Psychological Testing from a Behaviorist Standpoint," Greenspoon and Gersten (1967) lament the fact that behavioristic treatment procedures tend to be unrelated to the information obtained from psychological tests. They soberly point out that if the information obtained from psychological tests cannot be used to provide information about behavioristic treatment procedures, "then psychologists may have little reason to continue using the tests" (p. 849.).

child's social and physical environment in order to systematically evaluate the effects of various kinds of stimulus contingencies on the child's behavior.

On the other hand, "it may be possible to determine sources of control of the patient's behavior outside of the test situation from his performance on psychological tests" (Greenspoon and Gersten, 1968, p. 849). While the one-to-one testing session may lack many equivalences to the natural environment, the psychological examiner can administer tests under conditions which at least approximate certain aspects of treatment alternatives.[3] And, more importantly, he has the opportunity to systematically alter the stimuli confronting the child--in a single session--and look for functional changes in the child's behavior. For example, recognizing that verbal stimuli may have powerful A-R-D value--that present behavior may come under strong control of language--the clinical psychologist may wish to use an assessment session to determine how a variety of verbal stimuli reinforce and control the tested child's behavior. Such observations could serve as a partial basis for formulating a treatment plan and in making recommendations to others on how to best "relate" to the child. This kind

---

[3]Greenspoon and Gerstens's suggestions as to how the test situation should be structured in order to determine natural setting sources of behavior control are somewhat different than those offered here.

of assessment would be based on the recognized importance of situational determinants of behavior and on the assumption that the sample of behavior emitted on a test is related to behavior emitted by the test-taker in other settings which are in some ways equivalent to the test situation (stimulus generalization). Thus the testing situation might be considered as a microcosm of all learning-training situations; on the basis of the child's test behavior, the examiner may be able to make relatively precise statements about deficits or inappropriacies in a child's A-R-D system, make specific recommendations about the course and sequence of treatment designed to ameliorate A-R-D defects and instrumental behavior deficits, and also make predictions about how the child is likely to function in different kinds of training climates.

Interestingly a sine qua non of professionals in the testing field is that the psychological examiner create an interpersonal climate of warmth, acceptance and unconditional positive regard so as to elicit a child's "best" performance. Student examiners are admonished not to be harsh or critical toward their examinees. While such an interpersonal strategy may indeed increase the likelihood that the examiner will elicit from the child "what he knows," the employment of a homogeneous interpersonal manner throughout the testing session makes it difficult if not impossible to tell how the child would have performed if he had been criticized or had

been given no social feedback. Furthermore, the "warm" testing environment may have no resemblance whatsoever to the child's present training environment where he is experiencing problems.

While one objective of assessment is to determine what a child knows or has achieved or "where he's at" now, the second more important objective is to make appropriate recommendations about the training conditions under which the child can maximally function so as to acquire new academic and social skills (and, from the present point of view also so as to develop an adequate A-R-D system). Unfortunately, the use of assessment as a vehicle for making such precise training prescriptions has in the past taken a backseat to elaborate "diagnostic" descriptions of the tester's "current level of functioning" and the psychodynamic "causes" of his dysfunctions.

Disregarding for the moment the content of the specific tests to be employed, the following represents the outlines of a suggested assessment format--which is directly relevant to an A-R-D learning-principles conceptualization of behavior disorders of children and their remediation. The test battery would be divided into two equivalent halves. During administration of the first half, the child would not receive incentives during administration of the second half, incentives (e.g., money or edibles or tokens redeemable for a variety of "prizes") would be dispensed contingent upon

"correct" or task-appropriate responses. In addition, during the administration of each of the equivalent test batteries, the clinician would systematically alter his verbal behavior toward the child. For approximately one-third of the time, the clinician would withhold all social-evaluative verbal stimuli. For the second third of time, he would mildly criticize the child on an "incorrect"-response contingent basis. And during the last third, the clinician would praise the child on a "correct" response contingent basis. In a sense, then, the procedure would allow the clinician to observe the child's performance under six different, simulated, testing-training conditions--as shown in Table 10. In certain respects the assessment procedure might even be considered as a single-subject, clinical experiment. A hypothetical case will now be described in order to illustrate the applicability of the suggested assessment model and to dramatize the possible implications of its use.

Table 10

Testing Conditions in Proposed Child Assessment Model

	No Social Feedback	Criticism	Praise
No Tangible Incentives			
Tangible Incentives			

A nine year old, third grade boy was referred for psychological evaluation because of "shy" and "passive-aggressive" behavior and suspected learning

disability. In particular, the child's teacher observed that he avoided reading "at all costs" and was even seen to cry when asked to read aloud.

The child was administered the Stanford-Binet Intelligence Test and two alternate forms of a comprehensive reading test.[4] On the Stanford-Binet, items 1, 3, 5 and the alternate were administered in a standard manner, as was one of the reading tests. Then, items 2, 4 and 6 of the Stanford-Binet and the second reading test were administered --during which time the child was given pennies on a correct-response contingent basis.

The following hypothetical results were obtained. The child's overall IQ was measured at 115--his calculated score being somewhat higher on the "tangible reward" half of the test. On the reading tests, the child performed slightly below age level on the comprehensive portions but fell way down on the oral reading portions. Specifically, it was observed that the child had learned more than one dominant response to several simple words. For example, the child consistently made numerous discrimination errors between words such as "a--the," "here--were," "is--was," "gate-fence," etc.

In general the child's "attending" and "persevering" behaviors were far superior on the incentive phase of testing compared to the no-incentive phase. However, this statement required

---

[4]While clinical psychologists have traditionally and voluntarily restricted themselves to the use of "psychological" rather than "educational" tests, the exclusion of educational assessment devices should not be arbitrary; indeed, educational measures may occasionally provide the learning-oriented clinician with information as to "why" a child isn't learning or is misbehaving in the classroom. An A-R-D analysis is concerned with a child's skills or instrumental behavior repertoire as well as his A-R-D system since they have been clearly shown to interact. Furthermore it is suggested strongly that the tests which the psychologist employs should not be fixed--for example, always the WISC or always the Rorshach--but, rather, should be selected on the basis of relevance to the referral problem.

several qualifications. First, the child exhibited less task-oriented behavior during the administration of the reading test than the IQ test, and the number of his mislabeling responses was essentially the same under both the incentive and no-incentive phases of testing. During the incentive portion of testing, the child exhibited a high level of appropriate "striving" behaviors under both "praise" and "no feedback" conditions; under the "criticism" condition, however, while continuing to "strive," the child also emitted several self-depreciatory remarks (e.g., "I'm a dummy") and began "fidgeting" in his seat.

During the no-incentive phase of testing, the child was indiscriminately "distractible" under both praise and no-feedback conditions--emitting responses such as "I'm tired," "Is this all?" and "Can I go home now?" During the "criticism" condition in this phase of testing, the child became extremely "restless" and "passively oppositional"--to the point where he would not attend to the test stimuli nor verbally respond to the examiner.

The assessment session led to the following "diagnostic" conclusions:

(1) The child did have a learning disability with respect to reading. He had learned several bad habits (stimulus and response over-generalizations) which were interferring with his acquisition of new reading skills.

(2) The child also had a "motivational disability" or A-R-D defect with respect to reading, which apparently had been generalizing to other subject areas. That is, on the basis of his teacher's observations and his performance on the reading tests, reading stimuli apparently elicited strong negative emotional-avoidance responses.

(3) In view of the child's high level of "persevering" and "attending" behaviors during the incentive phase of testing organic interpretations of "hyperactivity" and "short attention span" were ruled out.

(4) Verbal praise was apparently not functioning as a strongly positive A-R-D stimulus for this child since no behavioral differences were noted between the "praise" and "no-feedback" testing conditions.

(5) "Criticism" stimuli had acquired strongly negative A-R-D value for this child. Not only did criticism stimuli serve to disrupt ongoing behavior, but also acted discriminatively to bring on a variety of maladaptive responses.

On the basis of these diagnostic conclusions, the following hypothetical treatment recommendations were made:

(1) During reading period each day, the child should be sent to the school's resource room where--under a token reinforcement system--he should be given individualized discrimination training on all words to which he responded both appropriately and inappropriately. In addition, the environment should be programmed so as to give the child ample opportunity to experience "success" under oral reading conditions.

(2) Since verbal praise had apparently not acquired strong emotional-reward value for this child, the resource room teacher would be instructed to pair various praise behaviors with the dispensing of tokens and tangibles on a heavily continuous reinforcement schedule. Later, in order to maximize resistance to extinction, the teacher would be advised to gradually reduce the frequency of praise and to administer social reinforcement on a variable ratio and interval basis.

(3) After further exploration of the child's social history and in view of his high IQ, it was decided to desensitize him to criticism-rejection (Wolpe, 1969).

(4) At the beginning of treatment, the child's homeroom teacher was advised to "lay off" him by not making any demands upon him and criticizing and disciplining him at a very minimum. As the child neared the top of the desensitization hierarchy, the homeroom teacher would be instructed to start phasing in increasing expectations and occasionally, but discreetly, to provide him with "constructive criticism." Such advisement was planned so as to be able to assess the effects of the desensitization and so as to enhance generalization to other "real world" situations in which the child would no doubt be finding himself in the future.

REFERENCES

Annon, J. S. The extension of learning principles to the analysis and treatment of sexual problems. Unpub. doctoral dissert., UH, 1971.

Anastasi, A. Psychological Testing. (3rd ed.) New York: MacMillan, 1968.

Atkinson, J. W. Motivational determinants of risk-taking behavior. In J. W. Atkinson & N. T. Feather (Eds.). A theory of achievement motivation. New York: Wiley, 1966, 11-29.

Atkinson, J. W. & Feather, N. T. A theory of achievement motivation. New York: Wiley, 1966.

Atkinson, J. W., & Litwin, G. H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. Journal of Abnormal and Social Psychology. 1960, 60, 52-63.

Atkinson, J. W., & O'Connor, P. Neglected factors in studies of achievement-oriented performance: Social approval as incentive and performance decrement. In J. W. Atkinson & N. T. Feather (Eds.). A theory of achievement motivation. New York: Wiley, 1966, 229-325.

Azrin, N. H. Time-out from positive reinforcement. Science, 1961, 133, 382-383.

Bandura, A. Influence of models' reinforcement contingencies on the acquisition of imitative responses. Journal of Personality and Social Psychology, 1965, 1, 589-595.

Barsch, R. H. Perspectives on learning disabilities: The vectors of a new convergence. Journal of Learning Disabilities, 1968, 1, 10-20.

Bateman, B. Learning disorders. Review of Educational Research, 1966, 36, 93-119.

Benton, A. L. Influence of incentives upon intelligence test scores of school children. Journal of Genetic Psychology, 1936, 49, 494-496.

Bijou, S. W., Peterson, R. F., and Ault, M. H. A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. Journal of Applied Behavior Analysis, 1968, 1, 175-191.

Birnbrauer, J. S., Wolf, M. M., Kidder, J. D., & Tague, C. E. Classroom behavior of retarded pupils with token reinforcement. Journal of Experimental Child Psychology, 1965, 2, 219-235.

Bornstein, A. V. The effects of examiner approval upon the performance of subjects on the performance scale of the Wechsler Intelligence Scale for Children. Unpublished doctoral dissertation, Florida State University, 1967.

Burt, C., & Williams, E. L. The influence of motivation on the results of intelligence tests. British Journal of Statistical Psychology, 1962, 15, 127-136.

Cohen, H. L. Educational therapy: The design of learning environments. In J. M. Shlien (Ed.), Research in Psychotherapy. Washington, D.C.: American Psychological Association, 1968, 21-53.

Cronbach, L. J. Essentials of Psychological Testing (2nd ed.). New York: Harper & Row, 1960.

Dinsmoor, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. Journal of Experimental Psychology, 1950, 40, 458-472.

Douvan, E. Social status and success striving. Journal of Abnormal and Social Psychology, 1956, 52, 219-223.

Exner, J. E. Variations in WISC performance as influenced by differences in pretest rapport. Journal of General Psychology, 1966, 74, 299-306.

Ferguson, H. H. Incentives and an intelligence test. Australian Journal of Psychology and Philosophy, 1937, 15, 39-53.

Gallimore, R. Variations in the motivational antecedents of achievement among Hawaii's ethnic groups. Paper read at East-West Center Conference in Culture and Mental Health, Social Science Research Institute, Honolulu, Hawaii, 1969. In W. Lebra (ed.) Mental Health Research in Asia and the Pacific, Vol. 2, East-West Center Press, Honolulu, Hawaii.

Gallimore, R., & Howard, A. (Eds.) Studies in a Hawaiian community: Na Makamaka O Nanakuli. Pacific Anthropological Records No. 1. Honolulu: Department of Anthropology, B. P. Bishop Museum, 1968.

Greenspoon, J., & Gersten, C. D. A new look at psychological testing: Psychological testing from the point of view of a behaviorist. American Psychologist, 1967, 22 848-853.

Guilford, J. P. The Nature of Human Intelligence. New York: McGraw-Hill Co., 1967.

Haggard, E. A. Social status and intelligence: An experimental study of certain cultural determinants of measured intelligence. Genetic Psychology Monographs, 1954, 49, 141-186.

Hull, C. L. Principles of behavior. New York: Appleton, 1943.

Klinger, E. Modeling effects on achievement imagery. Journal of Personality and Social Psychology, 1967, 7, 49-62.

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

Kubany, E. S., Bloch, L. E. & Sloggett, B. B. The good behavior clock: A reinforcement time-out procedure for reducing disruptive classroom behavior. Journal of Behavior Therapy and Experimental Psychiatry, 1971, in press.

Kubany, E. S., Gallimore, R., & Buell, J. The effects of extrinsic factors on achievement-oriented behavior: A non-western case. Journal of Cross-Cultural Psychology, 1970, 1, 77-84. To be reprinted in Rosenblith, J. F., Allinsmith, W., & Williams, J. P. (eds.) The Causes of Behavior: Readings in Child Development and Educational Psychology. Boston, Mass.: Allyn and Bacon, Inc.

Kubany, E. S., & Sloggett, B. B. Learning disability or motivational disability? Journal of Learning Disabilities, 1971, in press.

Lovaas, O. I. Behavior therapy approach to treatment of childhood schizophrenia. In J. P. Hill (Ed.), Minnesota Symposium on Child Development. Minneapolis: University of Minnesota Press, 1967, 108-159.

Lowell, E. L. A methodological study of projectively measured achievement motivation. Unpublished master's thesis, Wesleyan University, 1950.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. The achievement motive. New York: Appleton-Century-Crofts, 1953.

McClelland, D. C., Clark, R. A., Roby, B. T., & Atkinson, J. W. The projective expression of needs. IV. The effect of the need for achievement on thematic apperception. Journal of Experimental Psychology, 1949, 39, 242-255.

Meichenbaum, D. H., Bowers, K. S., & Ross, R. R. A behavioral analysis of teacher expectancy effect. Journal of Personality & Social Psychology, 1969, 13, 306-316.

Mingione, A. D. Need for achievement in Negro and white children. Journal of Consulting Psychology, 1965, 108-111.

Mischel, W. Continuity and change in personality. American Psychologist, 1969, 24, 1012-1018.

Patterson, G. R. An application of conditioning techniques to the control of a hyperactive child. In Ullmann and L. Krasner (Eds.) Case studies in behavior modification. Holt, Rinehart & Winston: New York, 1965.

Rosen, B. C. The achievement syndrome. American Sociological Review, 1965, 21, 203-211.

Rosonthal, R., & Jacobson, L. Pygmalion in the classroom: Teacher expectation and pupil's intellectual development. New York: Holt, Rinehart & Winston, 1968.

Sloggett, B. B. Behavior modification of the underachieving rural Hawaiian: An experimental classroom. Pacific Anthropological Records No. 5. Honolulu: Department of Anthropology, B. P. Bishop Museum, 1969.

Sloggett, B. B., Use of group activities and team rewards to increase individual classroom productivity. Teaching exceptional children, 1971, 3, 54-66.

Sloggett, B. B., Gallimore, R., & Kubany, E. S. A comparative analysis of fantasy need achievement among high and low achieving male Hawaiian-Americans. Journal of Cross-Cultural Psychology, 1970, 1, 53-61. To be reprinted in Rosenblith, J. F., Allinsmith, W., & Williams, J. P. (eds.) The causes of behavior: Readings in child development and educational psychology. Boston, Mass.: Allyn and Bacon, Inc.

Staats, A. W. Learning, language and cognition. New York: Holt, Rinehart & Winston Inc., 1968.

Staats, A. W. Social behaviorism, human motivation, and the conditioning therapies. In B. A. Maher (Ed.), Progress in experimental personality research. New York: Academic Press, 1970.

Staats, A. W. Child learning, intelligence, and personality. New York: Harper and Row, 1971a.

Staats, A. W. Learning and personality, and language (symbolic) behavior therapy. Behavior Therapy, 1971b, in press.

Staats, A. W., Minke, K. A., Goodwin, W., & Landeen, J. Cognitive behavior modification: "motivated learning" reading treatment with subprofessional therapy-technicians. Behavior Research and Therapy, 1967, 5, 283-300.

Staats, A. W. (with contributions by Staats, C. K.) Complex Human Behavior. New York: Holt, Rinehart and Winston, Inc., 1963.

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

Thistlewaite, D. A critical review of latent learning and related experiments. Psychological Bulletin, 1951, 48, 97-129.

Tiber, N., & Kennedy, W. A. The effects of incentive on the intelligence test performance of different social groups. Journal of Consulting Psychology, 1964, 28, 187.

Tolman, E. C., & Honzik, C. H. Introduction and removal of reward, and maze performance in rats. 1930, University of California Publications in Psychology, 4, 257-275.

Wendt, H. W. Motivation, effort, and performance. In D. C. McClelland (Ed.) Studies in motivation, 1954, New York.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Winterbottom, M. R. The relation of childhood training in independence to achievement motivation. Unpublished doctoral dissertation, University of Michigan, 1953.

Wolpe, J. The practice of behavior therapy. New York: Pergammon Press, 1969.