1977

Effects of extrinsic rewards on the subsequent choice behavior of academically delayed children.

George Ramey
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/theses


This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
EFFECTS OF EXTRINSIC REWARDS ON THE SUBSEQUENT CHOICE BEHAVIOR OF ACADEMICALLY DELAYED CHILDREN

A Thesis Presented
By
Gregory Ramey

Submitted to the Graduate School of the University of Massachusetts in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
April 1977
Psychology
EFFECTS OF EXTRINSIC REWARDS ON THE SUBSEQUENT CHOICE BEHAVIOR OF ACADEMICALLY DELAYED CHILDREN

A Thesis Presented
By
Gregory Ramey

Approved as to style and content by:

Beth Sulzer-Azaroff, Chairperson of Committee

James M. Royer, Member

J. Gregory Olley, Member

Bonnie Strickland, Department Head
Psychology Department
In Appreciation

To Beth Sulzer-Azaroff, whose valuable help and support throughout this project have been immeasurable.

To James M. Royer and J. Gregory Olley, whose participation as members of my thesis committee has been greatly appreciated.

To the personnel and children at Ft. River School, Amherst, Massachusetts, whose cooperation made this project possible.

And to Marilyn Fischer, for being there.
# Table of Contents

List of Tables.........................................................v
List of Figures.........................................................vi
Abstract.................................................................vii
Introduction..............................................................1
Theoretical Perspectives...............................................5

## Literature Review

- Expected rewards contingent upon participation.............14
- Expected rewards contingent upon performance..............20
- Noncontingent unexpected rewards..............................24
- Verbal rewards.........................................................26

Methodological Concerns.............................................30

Method.................................................................36
Results.................................................................55
Discussion..............................................................70

References.............................................................83
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Type of reward and effect on intrinsic interest</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>Types of problems on each worksheet</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>Reward conditions associated with each experimental phase</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>Choice, rating, and on-task behavior of subjects 1-6</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>Problems attempted and accuracy during tutoring and choice periods</td>
<td>68</td>
</tr>
</tbody>
</table>
List of Figures

Figure
1 Rating chart used by subjects after completing each worksheet .......... 40
Percent of on-task behavior during tutoring. Worksheets selected during immediate and delayed choice periods
2 Subject 1.......................... 59
3 Subject 2.......................... 60
4 Subject 3.......................... 61
5 Subject 4.......................... 62
6 Subject 5.......................... 63
7 Subject 6.......................... 64
ABSTRACT

EFFECTS OF EXTRINSIC REWARDS ON THE SUBSEQUENT CHOICE BEHAVIOR OF ACADEMICALLY DELAYED CHILDREN

(April 1977)

Gregory Ramey, B. A., Lake Forest College
Ed. M., Harvard University
Directed by: Professor Beth Sulzer-Azaroff

Research examining the relation between extrinsic rewards and subsequent interest in an activity has generally found that when the extrinsic reward is no longer available, interest in the task is decreased from pre-reward levels. However, several important aspects of these studies, such as the failure to differentiate between reward and reinforcement procedures, nature of the experimental task, and others place critical limitations on the generalizability of these data.

The current investigation was undertaken to ascertain the relation between extrinsic rewards (token reward and verbal reward) and intrinsic interest in an activity (as defined by choice of activities and self-report of enjoyment) by following a procedure that more closely paralleled the conditions under which applied reward systems are typically instituted.
Six elementary school children, identified as needing extra assistance in math, were selected as subjects. A multielement design, in which different types of math problems were correlated simultaneously with different reward conditions, was utilized. Children completed three math worksheets during a daily tutoring session. One worksheet was associated with no reward—the child simply completed the sheet and returned it to his tutor. A second worksheet was correlated with a verbal reward, in which both general and specific praise comments were delivered contingently for accuracy of work. A third worksheet was correlated with a token reward system, in which points (redeemable for small toys) were given for each problem solved accurately. Intrinsic interest was assessed by having the child select one of the three worksheets immediately after the session, and again at a later time in the day. These two choice worksheets were completed under the no-reward condition. Subjects also rated their enjoyment of each worksheet. Various other parameters of performance (on-task behavior, accuracy, number of problems attempted) were also collected.

Three experimental phases were conducted. During the Baseline phase, the six subjects completed
all three types of worksheets under the no-reward condition. During the Token-Low Choice phase, subjects completed the low choice activity under the token condition, the medium choice activity under the verbal reward condition, while the high choice activity continued to be completed under the no-reward condition. Three subjects, because they exhibited a high preference for a particular activity, were placed in a Token-High Choice phase. These subjects continued to complete the medium choice activity under the verbal reward condition, but the low choice activity was switched to the no-reward condition. The high choice activity was placed in the token reward condition.

In general, children's interest in both high and low interest tasks, as measured by self-report and choice of activities, was enhanced rather than decreased by token rewards. The effect of verbal rewards was more ambiguous. The distinguishing characteristics of this study are noted in an attempt to explain the apparent discrepancy with previous research.
Introduction

It has been estimated that up to 20% or more of school-age children exhibit serious learning or behavioral characteristics that hinder significantly their educational or personal development (Gardner, 1977). A variety of intervention strategies have been developed to assist such children. Since the pioneering work of Staats, Staats, Schutz, and Wolf (1962) with problem children, and Ayllon and Azrin (1968) with adult psychiatric patients, behaviorally based intervention techniques have been employed with increasing frequency. One such treatment package, a token economy program, involves the systematic use of rewards on a frequent basis to increase appropriate behavior. While such a package has proved to be an extremely effective treatment approach for a variety of problem behaviors in both children and adults, the maintenance of change over an extended period of time, and generalization of responses to new environments has proven to be a more difficult task (Kazdin & Bootzin, 1972; O'Leary & Drabman, 1971).

There are at least two possible and not incompatible explanations for such difficulty. One is the absence or inadequate design of a systematic plan for the maintenance and generalization of the desired
behavior change. When present, such a plan may involve manipulating several potentially important factors. Reinforcers may gradually be delivered less frequently, on a more intermittent schedule. Individuals in the natural environment might be instructed in appropriate techniques for eliciting and consequating the desired behavior. Natural rewards (e.g., smiles, verbal praise, etc.) may be presented concomitantly with "artificial" or arbitrary rewards (e.g., candy, tokens, etc.) so that the former would eventually exert the same degree of control as the latter. Failure to maintain a particular behavior is not seen as a deficiency inherent in the reward structure, but rather as a failure to extend and modify that system to adapt to the contingencies of the natural environment.

A second possible explanation suggests that the same reward systems that are successful in achieving short-term desirable changes may unintentionally lead to different long-term negative effects. Since every delivery of a reward by others is inherently an act of communication, giving a reward for performing a behavior could communicate that the activity is not worth doing for its own sake. Thus, a person's intrinsic interest in the activity, behavior emitted
in the absence of any obvious reward structure, may be inadvertently decreased, since participation is more likely to become simply a means to an end (i.e., a reward) rather than an end in itself.

This concern with the potentially inimical effects of extrinsic rewards has recently spurred a proliferating number of studies. One review of this literature concluded that the available evidence suggests that such extrinsic rewards as tokens should be avoided "unless there is a real danger to the individual or there is no alternative" (Levine & Fasnacht, 1974, p. 820). The suggestion that systematic extrinsic rewards should be reserved as a last resort may have the effect of discouraging the use of a clinically proven effective technique of behavior change. Since the need for effective intervention strategies for behavior change remains great, it thus becomes important to specify under what conditions token systems may be used legitimately.

Following a brief summary of the major theoretical perspectives relating to this research, an overview of the results of these studies will be offered. In general, such studies have examined the effect of expected rewards contingent upon participation in an activity, expected rewards contingent upon performance,
unexpected rewards, and verbal rewards. Following a description of these studies, several potentially important methodological ambiguities will be examined—conflicting definitions of intrinsic interest, failure to differentiate between a single reward and a reinforcement procedure, verification of the understanding of expected reward conditions, magnitude of reward, duration between reward and measurement of interest, performance data and the nature of the experimental task.
Theoretical Perspectives

Much of the research examining the relation between extrinsic rewards and intrinsic interest falls within the framework of causal attribution theory. This theory has focused on both interpersonal and self-perception attributions. Interpersonal causal attribution concerns how a person makes inferences about the locus of causality for the actions of others (Kelly, 1973). The significance of this research is the assertion that such causal inferences can influence the subsequent behavior of the observer (Kopel & Arkowitz, 1975). For example, suppose a student brought an apple to his teacher every day. It is highly probable that the teacher's subsequent behavior towards the student would be influenced by the extent to which the teacher perceives the cause of the student's actions as motivated by kindness, fear, jealousy, or a desire to ingratiating.

A second kind of attribution theory concerns the self-perception of attribution. Self-attribution refers to inferences a person makes about the cause of his own behavior. Again, the import of this analysis is that such inferences may affect subsequent behavior. To illustrate, the role of self-attribution in the maintenance of behavior change was examined by Davison,
Tsujimoto, and Giaros (1973). Insomniacs were treated with a combination of relaxation and scheduling procedures. In addition, clients were also given identical doses of chloral hydrate. Half of the subjects were told that the drug should have a minimal effect on their sleeping patterns, while other clients were told that the drug should greatly facilitate their sleeping. During post-treatment follow-up, when the drug was no longer used, the clients who had been instructed that the drug was of minimal value reported greater therapeutic gains than the other group. The hypothesized explanation of these data is that the minimal group perceived behavior changes to be related to self-attribution, while the drug group perceived changes in sleeping to be related to external factors (i.e., chloral hydrate).

But how do individuals make such inferences? Two alternative possibilities have been suggested. Bem (1967, 1972) has delineated a radical behavioral self-perception theory that views inferences of causality as arising from an observation of one's own overt behavior within a given situational context. Such external cues, based on an individual's reinforcement history, provide the basis for self-perception judgements. Having labelled or inferred a cognition
or internal state based on an observation of one's own overt behavior, such cognitions can influence subsequent behavior. Suppose a student continues to solve math problems during study time even though he completed his assignment. The student, having "observed himself" working on math in the absence of any obvious external incentive system, may infer that he must enjoy doing such work. This inference, based on the person's self-perception of his own behavior within a situational context, may then have important implications for future behavior.*

*Can the use of such inferences be justified within a radical behavioral framework? Mahoney (1974) has argued quite convincingly that it can, although "right-wing behaviorism" has traditionally regarded inferences as "technical obscenities" to be supplemented by a vow of "strict inferential celibacy" in pseudo-imitation of the physical sciences. This basic misunderstanding of Behaviorism ignores the fact that inferences are an inevitable component of any scientific endeavor. Indeed, many critical constructs of the physical sciences are hypothetical, inferred phenomena (e.g., an electron). "Our current concern is not whether inferences are justified, but rather when and which
An alternative formulation of Bem's self-perception theory is Deci's cognitive evaluation theory (Deci, 1975a, 1975b). This orientation places primary emphasis on "self-knowledge" as a relevant factor. According to Deci, there are two cognitive processes through which rewards may affect self-attribution, and subsequent intrinsic interest. The first component is the controlling aspect of rewards. That is, rewards are frequently used to get a person to do something that ordinarily he may not do—i.e., work at a job. This controlling aspect of rewards communicates to the person that the task is not worth doing for its own sake. All rewards, contends Deci, have this component, even rewards given for high interest activities. For example, if a person

inferences add to our understanding of behavior...

an inference is justified if, and only if it increases predictive accuracy or conceptual breadth. To the extent that a presumed element can be shown to be useful in predicting, controlling, or understanding systematic relationships, then it is logically warranted” (Mahoney, 1974, p. 30, 32).
enjoys playing with marbles, the introduction of payment for doing the task may communicate that the activity is not worth doing without pay. Thus, the person may come to reattribute the cause for his actions as external to himself (receiving money), rather than internal (inner satisfaction). Money, an extrinsic reward, has thus caused a phenomenological transformation in the interpretation of the task, with a resulting undermining of the person's initial intrinsic interest to play "for its own sake." Thus, this first aspect of rewards has altered the perceived locus of causality of the person's behavior.

But cognitive evaluation theory suggests that there is a second aspect of rewards—a feedback function, conveying information affecting a person's feelings of competence and self-determination. A reward communicates that the person is doing something correctly, that the activity is valued. When this feedback function of a reward is high, a person is more likely to perform the activity in the future. But both components are inherent in any reward. Which process is dominant—feedback or controlling—depends upon which aspect of the reward is more salient. Money or tokens may tend to decrease intrinsic interest because such rewards have a high
controlling feature, due to their past association with work activities. However, such rewards as praise or social approval may tend to increase intrinsic interest, because those rewards have more of a feedback function, increasing a person's sense of mastery over the environment.

What is the difference between Bem's behaviorally oriented and Deci's cognitive explanation of self-perception and its subsequent effects? According to Deci (1975b, p. 285), the difference is as follows:

\[ I \text{ anything...}\text{asserting that people have personal knowledge of their own internal states which are knowledgeable only to them (unless they report it), and that they make attributions to others largely through knowing what their own internal state would probably be if they were in the position of the actor. Bem's position on the other hand asserts that people make attributions to others and to themselves by observing external cues and then inferring their internal states from their observations.} \]

In fact, this alleged distinction is somewhat artificial, since Deci does acknowledge the importance of external forces in influencing "personal knowledge."
Correspondingly, Bem acknowledges that private events or cognitions can influence future behavior.

The difference between orientations is apparently focused on the origin and emphasis of the role of private events or personal knowledge. Neither formulation denies that private events exert an important influence on future behavior, although the cognitive orientation does place a greater emphasis on the "internal" origin of such knowledge. In contrast, Bem's behavioral explanation denies any primacy or special status to internal states. That is, the same functional relation used to explain overt behavior can be extended to an analysis of cognitions or covert behavior. In apparently ascribing to personal knowledge some special status, Deci is suggesting some element of discontinuity between laws regulating overt and covert behavior.

In addition to the apparent disagreement concerning the origin of private events, there is also some dispute regarding the role of inferred processes. Insofar as such processes are unobservable, the invoking of such constructs by radical behaviorists requires a careful justification that their utility is empirically justified--i.e., contributes to an explanation, and not merely a pseudo-description of
the phenomena. Since a scientific explanation requires relating a phenomenon to some law of behavior, prediction or experimental control should be facilitated by such constructs. The difference between orientations, with respect to the role of inferred processes, thus seems one of degree rather than qualitative distinctions.

It is difficult to envision a critical experiment that would differentiate between cognitive and behavioral self-attribution theories. Rather, the ability of such frameworks to predict human behavior will be the ultimate test of either orientation. But irrespective of such difficulty, the attribution orientation has spurred a number of studies attempting to define how extrinsic rewards may affect intrinsic interest in a task. Such experiments have generally been placed into four categories, depending on the type of reward system used: expected rewards contingent upon participation, expected rewards contingent upon performance, unexpected rewards, and verbal rewards. However, before reviewing these studies, some definitions of intrinsic interest will be offered.

**What is intrinsic interest?**

Intrinsic interest is defined as behavior emitted
in the absence of any apparent reward structure, behavior for which the only obvious reward is the activity itself. A variety of conceptualizations has been offered to explicate this rather imprecise definition (Deci, 1975b, p. 23-62). Some (Deci, 1975b, p. 61) have viewed such behavior as arising from "intrinsic motivation," and developing from a need of people to "feel competent and self-determining." Others have viewed intrinsic motivation or interest as developing from a need for uncertainty reduction (Festinger, 1957), optimal incongruity (Hunt, 1965), or a desire to fulfill an exploratory (Montgomery, 1955) or manipulation (Harlow, 1953) drive. "Intrinsic interest" as used in this paper is a generic term, encompassing various conceptualizations of such behavior. As will be noted later, however, this array of definitions of intrinsic interest or motivation has led to several important difficulties.
Literature Review

Expected rewards contingent upon participation

The largest number of studies has dealt with the delivery of an expected reward not contingent upon any specific response, but rather contingent only on participation in the experimental sessions. That is, all subjects within a given group were given the same type and magnitude of reward based simply on their participation, and not on achieving any specified criteria. Rewards have included such items as money, promise of experimental credit, prizes, certificates, food, and participation in desired activities. The reward was expected insofar as the subjects were told before the experimental session what they were to receive.

Weick's study (1964), using 100 male college students as subjects, involved two groups—one of which was given experimental credit for participation in a concept attainment task, while the other was not. The group that was denied credit persisted longer on the task, performed better, and rated the session as more interesting. In a similar experiment with 32 Israeli high school students, Kruglanski, Friedman,
and Zeevi (1971) promised half of the subjects a guided tour of the psychology department as a reward, while no mention of a reward was made to the other group. In tasks involving recall and creativity, subjects who were not promised an incentive performed better, and rated the task as more enjoyable.

Lepper, Greene, and Nisbett (1973) examined the effect of a token reward on children's drawing activity. Fifty-one subjects, ranging in age from 40 to 60 months, were selected on the basis of their initial high interest in a drawing activity, as reflected by the amount of time they spent using the magic markers during class time. The children were individually taken into a room, and urged to draw a picture with a magic marker. One group of children was promised and given a 'Good Player Award' for their drawings, while the other group was given no tangible reward. Seven to fourteen days after the reward session, measures were taken of the children's interest in the drawing activity. This measure was compared with pre-experimental baseline measures of their drawing behavior in class. The children in the rewarded group displayed significantly less interest in the drawing activities than the non-rewarded group. The quality of the drawings for the group given rewards
was also judged significantly inferior. This same finding was replicated by Greene and Lepper (1974) in two different studies. The first study was a replication of Lepper et al. (1973). The second part of the study used the same format, but with puzzles instead of magic markers, and access to attractive toys as a reward instead of a token reward. In both instances, it was the rewarded group that displayed less subsequent interest in the activity, as measured by post-experimental behavioral measures of time spent on the task during class time.

Reiss and Sushinsky (1975), using 32 first grade girls aged 6 to 7 years as subjects, found the same inverse relationship between the dispensing of a reward (access to playing with a doll), and subsequent interest in a specified target behavior (listening to music).

With behaviors of high frequency or interest, the previously discussed studies have consistently found that the giving of a reward was correlated with a decrease in later interest in the activity. However, Calder and Staw (1975b) examined the effect on both high interest (attractive picture puzzle) and low interest (blank puzzle) tasks. Forty under-
graduate males served as subjects in an experiment in which half of the group was promised and given one dollar for solving 15 puzzles. Half of the subjects were asked to complete attractive puzzles, while the others were given the blank puzzles. Two measures of intrinsic interest were taken—reported task satisfaction, and willingness to volunteer for future experiments without pay. A disordinal interaction between puzzle interest and money was substantiated. That is, with the blank puzzle (low interest activity), money increased the reported interest in the task while the lack of payment decreased interest in the task. As in other studies, however, money decreased interest in the high interest activity (attractive puzzle), while lack of money increased interest in the puzzle. This finding was important, for it suggested that noncontingent expected rewards decrease intrinsic interest only in tasks that were initially of high interest.

A recent study by Kruglanski, Riter, Amitai, Margolin, Shabtai, and Zaksh (1975b) suggested an additional attribute to rewards——whether they are peripheral or integral to the task. The experiment involved 48 Israeli boys between the ages of 14 to 15, with intrinsic interest being assessed by a
post-experimental questionnaire. Two experimental tasks were used. In a money tossing task, it was assumed that since the game is always played with money, payment would be perceived as an integral aspect of the activity. Hence, the reward (money) should not decrease intrinsic interest. The second experimental task involved the use of wooden bricks to build a model, an activity that is not ordinarily associated with payment. Hence, it was hypothesized that money would be perceived as an extrinsic feature of the game, with a resulting decrease in intrinsic interest. The hypotheses were confirmed. The reward decreased intrinsic interest only in the task (i.e., building blocks) in which money was regarded as extraneous. In a second part of the experiment, this result was replicated with two different tasks, one in which money was an integral part of the activity (Stock Market game), and another in which it was peripheral (Athletics game).

Using a group of mentally retarded adolescents as subjects, Lee, Syrnyk, and Hallschmid (1977) assessed the effects of individually tailored rewards on both a high and low interest task. Subjects' preference for six incentive objects (coke, card, nail clipper, pencil, comb, and paper clip) was first
assessed by asking the subject which of two objects he would like the most. Each incentive was paired with every other, thus resulting in an individual reward hierarchy (based on verbal report of preference) for each child. Half of the subjects were rewarded with their high-incentive reward, while half were rewarded with their low-incentive reward. The task involved playing with either a high interest (attractive xylophone) or low interest (unattractive xylophone) task. Subsequent interest was assessed by the time spent playing with the instrument during free time. For the high interest activity, task persistence was greater for the low reward than for the high reward condition. For the low interest activity, task persistence was greater for the high reward than for the low reward condition.

How would self-perception theory account for these data? Behavior (e.g., solving puzzles, drawing, etc.) performed in the presence of an extrinsic reward generally leads to the individual's inference that his behavior was caused by the reward structure. Thus, when the reward is removed, the interest in the activity is decreased. Within Deci's cognitive evaluation theory, rewards such as money generally have more of a controlling rather than an informative feature,
usually serving to decrease interest in the activity. Thus, the studies reviewed to date have found that rewards contingent upon participation are associated with: (a) a decrease in interest in high probability behaviors; (b) an increase in interest in low probability behaviors; and (c) an increase in interest if the reward is regarded as an integral part of the task (e.g., coin tossing). The one exception to these generalizations is a study reported by Deci (1972b) in which intrinsic interest, as measured by free time spent on solving puzzles, was not affected by a monetary reward.

**Expected rewards contingent upon performance**

Expected rewards are delivered contingent upon performance when the rewards are given conditionally not only upon participation in the experiment, but also upon performing at a specified criterion level. That is, subjects are differentially rewarded dependent upon how many times they manifest a specific behavior.

Deci (1971) has reported two different studies that deal with the effect of contingent expected rewards on intrinsic interest. In the control group, subjects worked on Soma puzzles without pay
for three sessions. During session two, subjects in the experimental group were paid one dollar for each correct solution to the puzzles. In the middle of each session, the experimenter left the room for eight minutes. The subject was told he was free to do anything he wished during that time. The amount of free time spent working on the puzzles in the experimenter's absence was taken as a behavioral measure of intrinsic interest. The group that did not get paid persisted longer on the puzzles during their free time, suggesting that contingent rewards also decrease intrinsic interest. However, there was no difference in performance within each session between the experimental and control groups. These results were replicated by Deci (1972a). A second aspect of the 1971 study involved 8 students who worked as headline writers for the college newspaper. The index of intrinsic interest was the amount of time it took to write each headline, the assumption being that the more highly motivated would write more quickly. Again, it was the group that received payment that manifested a lower index of intrinsic interest (i.e., they took longer to write the headlines).

Working in a more naturalistic setting with
30 pre-school children ranging in age from 48 to 63 months, Lepper and Greene (1975) rewarded the correct solution of a series of puzzles with the opportunity to play with highly attractive toys. Free time spent playing with the puzzles during class was the measure of intrinsic interest. It was the rewarded group that manifested decreased interest in the activity after the experimental sessions.

In another study, Greene, Sternberg, and Lepper (1976) examined the effects of the introduction of a token economy reward system in increasing both low and high frequency behaviors in a math resource room. Forty-four fourth and fifth grade students were randomly assigned to one of four groups: The high interest group was rewarded for activities that they had selected most frequently during a 13 day baseline period. The low interest group was rewarded for activities that they had selected least frequently. In the choice group, the children chose which two activities they wanted to be rewarded for, while a control group received rewards for participation in all 4 activities. The amount of time spent interacting with specified math activities (but not performance on that activity) was rewarded. Following a 13 day treatment phase, the reward system was abruptly terminated, with data
collection continuing for 13 more days. In a within-group comparison, both the high-interest and choice groups spent significantly less time on the activities during the withdrawal period than they did during baseline. In comparison with the matched control group (between groups analysis), students in the low-interest and choice groups spent significantly less time on the activity, while no significant difference was noted in the high-interest group.

The studies examined so far have suggested that expected rewards contingent upon performance also result in a decrease of intrinsic interest once the reward is removed. The one exception to this generalization was a study reported by Reiss and Sushinsky (1975). Nine kindergarten children served as subjects for discrimination training in the correct identification of songs. Correct responses were rewarded by poker chips, redeemable for attractive toys. The children's subsequent interest in the songs, rather than being undermined by receiving rewards for listening to them, was in fact enhanced. The amount of free time they spent in class listening to the rewarded songs was significantly greater than the time spent listening to non-rewarded songs.
As with rewards contingent upon participation, self-perception theory focuses attention on the fact that behavior performed in the presence of a salient (i.e., obvious or obtrusive) reward leads a person to conclude that he is probably performing the activity simply for the reward. Thus, when the reward is unavailable, interest will decrease.

**Noncontingent unexpected reward**

In these series of studies, subjects were not told until after the experimental sessions were over that they were to get a special reward. What effect did this unexpected reward have on their subsequent interest in the activity?

Several studies have consistently shown that when the reward is unexpected, there is no subsequent loss in interest in the activity. Lepper, Greene, and Nisbett (1973) and Greene and Lepper (1974) reported studies of preschool subjects who initially displayed high interest in drawing with felt-tip pens. The experimental sessions consisted of having the child draw a picture, and then, unexpectedly, receive a certificate for his work. Seven to fourteen days after the unexpected reward session,
the amount of time spent voluntarily drawing in class was assessed and compared with baseline measures. The children who were given an unexpected reward, as well as those children who were not given a reward, displayed no change in interest in the activity. Lepper and Greene (1975) replicated these results with a task of solving puzzles, and a reward of playing with attractive toys.

The only study reporting a decrease in intrinsic interest as a function of unexpected rewards was that of Kruglanski, Alon, and Lewis (1972). The experiment involved 69 Israeli children 10 to 11 years of age. In the experimental group, members of the winning team were given an unexpected prize for winning a series of team competitions, while a control group was not given any prize. Intrinsic interest was assessed by subjective ratings of task satisfaction both immediately after the task, and one week later. In both instances, it was the group given the prize that reported less enjoyment of the competitions. This effect has not been replicated in any of the studies under review.

In explaining these results, attribution theory would suggest that since the person was not performing the activity for some extrinsic reason, the con-
trolling aspect of the reward would be minimal, while the feedback aspect would take on greater importance. Hence, there would be no subsequent loss of interest in the activity.

**Verbal reward**

The evidence to date has suggested that contingent rewards decrease interest, while unexpected rewards given after the task do not have any effect on subsequent interest. But since many rewards in everyday life take the form of verbal feedback, an intriguing question is what effect such verbal statements have on intrinsic interest.

**Noncontingent positive verbal reward.** Deci (1971) has reported some marginally significant results suggesting that positive verbal feedback may increase intrinsic interest. Twenty-four subjects were asked to solve puzzles in three different sessions, with intrinsic interest being measured by the amount of time the subjects would play voluntarily with the puzzles when the experimenter left the room for 8 minutes in the middle of each session. But instead of receiving money, the experimental group was praised for their performance at the end of each puzzle. The group that was given such feedback
displayed more subsequent interest in the puzzle during free time.

This result was refined in two further studies. Deci (1972a), using a similar methodology to that reported above, replicated the results for males only. Paradoxically, there was an opposite effect for females—-intrinsic interest tended to decrease as a function of verbal feedback, although this tendency was not statistically significant. However, a replication of that experiment (Deci, 1975b) using both male and female experimenters did find a significant male-female difference. Positive verbal feedback during the experimental sessions increased subsequent interest in the activity for males, while it decreased interest for females.

**Positive verbal feedback and tangible reward.** Only one study has been reported in this area, with Deci (1972b) suggesting that when expected tangible rewards (which tend to decrease intrinsic interest) and positive verbal feedback (which tends to increase interest) are combined, there is no subsequent effect on interest in the materials. Again, in this experiment, the free choice paradigm was used as the measure of intrinsic interest.
In attempting to account for the apparent discrepancy between results for males and females, Deci (1972a) has suggested an inverted U relationship between strength of verbal rewards and its subsequent effect on interest. It may be that such interest increases with increased verbal feedback only until a certain point, when it begins to decrease. With an attribution framework, this relationship would be explained in terms of a change in the function of verbal praise from feedback to controlling. While some praise may increase one's feeling of competence and mastery, increased amounts may lead the person to believe that he is being controlled or manipulated by others. If socialization patterns result in females becoming more dependent on verbal feedback than males, then the same amount of verbal praise may exert a more powerful influence on females, with a resulting change in the component of the reward as more manipulative.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lepper et al.</td>
<td>1974</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creeme et al.</td>
<td>1974</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepper et al.</td>
<td>1973</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kriegel et al.</td>
<td>1972</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reses et al.</td>
<td>1975</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee et al.</td>
<td>1977</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kriegel et al.</td>
<td>1975</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calder et al.</td>
<td>1975</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reses et al.</td>
<td>1975</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calder et al.</td>
<td>1975</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Type of Reward and effect on intrinsic interest.
Methodological Concerns

As noted in Table 1, there are some definite trends in the research previously reviewed. However, there are some critical questions concerning the interpretation and generalizability of these data.

1. What is intrinsic interest? A fundamental problem in reviewing studies on intrinsic interest and rewards is the wide variety of operational definitions. Three definitions of intrinsic interest have generally been offered. (a) Task satisfaction. In studies such as Kruglanski et al. (1972), intrinsic interest was inferred from a rating of how well the subjects enjoyed the experimental task. (b) Task persistence. In the studies of Deci (1972a, 1975a), and Greene et al. (1974), the amount of free time the subjects interacted with the task when not required to do so was the measure of intrinsic interest. (c) Task performance. In Kruglanski et al. (1971), and Weick (1964), the performance of the subjects during the experiment was considered an aspect of their intrinsic interest. That is, higher levels of interest were inferred from better performances.

The difficulty in this area is not that various measures of intrinsic interest have been used. Rather,
the problem is in the assumption that such disparate measures as verbal report of enjoyment, task persistence, and choice are all members of the same response class, and vary together. Thus, to avoid both semantic confusions and theoretical ambiguities, it would seem more valuable to relate specific independent variables (e.g., money, verbal feedback, awards) with specific dependent measures. Only as evidence regarding the covariance of these measures is accumulated would more global assertions of the relationship between intrinsic interest and extrinsic rewards be warranted.

2. **Reward versus reinforcement.** There has frequently been a failure to differentiate clearly between the delivery of a single reward and a reinforcement operation. Reinforcement refers to a procedure whereby "the contingent use of a stimulus results in an increase or maintenance of a dependent behavior" (Sulzer & Mayer, 1972, p. 293). To discuss the effect of extrinsic reinforcement on intrinsic interest demands (a) an initial baseline measure of a target behavior and (b) a measurement of the target behavior during delivery of the reward to ascertain whether reinforcement (maintenance or increase of behavior) has, in fact, taken place. For example, the Lepper et al. (1973)
study involved the delivery of a reward to children after they had completed a drawing activity. The drawings of the children who were given a reward were judged qualitatively inferior to those of the nonrewarded group. While a reward was given in this experiment, this was clearly not a reinforcement procedure.

The question arises as to the relevance of the reviewed studies in making generalizations about the applied use of rewards in systematic reinforcement procedures (e.g., token economy systems). That "tokens tend to decrease the intrinsic value of an activity" (Levine & Fasnacht, 1974, p. 819) is an unwarranted extrapolation from the available evidence.

3. Verification of understanding of conditions. Only in the Reiss (1975) study was the issue of verifying the expectedness of the reward actually confirmed. In Greene and Lepper (1974), and Lepper et al. (1975), the experimenters were dealing with very young children. Yet, no evidence was offered that the children actually understood the nature of the contingency. Reiss (1975) controlled for this factor by not proceeding with the experiment until the subjects could correctly answer questions about the reward contingencies.
4. **Magnitude of reward.** While much attention has been focused on the nature of the rewards (e.g., money, verbal feedback, etc.) the magnitude of the reward has not been considered. Yet, this may be a crucial variable in the self-perception process. This can most clearly be seen with verbal feedback. While the evidence suggests that verbal feedback can increase intrinsic interest in males, increasing the magnitude of the reward may result in the feedback being perceived as ingratiating, with a resulting decrease in intrinsic interest. A similar effect may occur with unexpected rewards.

5. **Duration between reward and measurement of interest.** While there may be a relation between rewards and intrinsic interest, the magnitude and duration of such an effect remains unclear. Many studies have conducted measurement of interest during or immediately after the reward session. Yet, such effects may be quite transitory. Kruglanski et al. (1972) took measures not only immediately after the task, but also one week later. Greene et al. (1976) continued 13 days of monitoring after the reward was removed. Periodic measurements over a longer period of time would yield valuable information about the stability of any effect.
6. **Performance data.** The question arises as to whether decreased subsequent interest in a rewarded task is the result of increased effort that was expended during the experimental session. Performance data reported by Deci (1975a) suggested that the amount of time spent by the subjects in solving the puzzles was approximately equal between the rewarded and nonrewarded groups. However, it would be a mistake to equate time spent on an activity with effort expended. Satiation or fatigue factors are not directly related to temporal involvement. This factor could be better controlled by increasing the amount of time between the reward session, and the subsequent measures of interest in the activity.

7. **Nature of experimental tasks.** A potential limitation to the external validity of the studies under review concerns the nature of the experimental tasks. This is particularly important in looking at the effect of rewards on low interest behaviors. Rather than decreasing interest in low frequency behaviors, rewards may serve to stimulate interest in an activity for which the person has not interacted successfully. Even in the study that most clearly approximated a token economy situation (Greene et al., 1976), the experimental task involved increasing time spent with
the activity, and not accuracy or performance.

Rationale for current research

The current investigation was undertaken to ascertain the relation between extrinsic rewards (verbal praise and token reward) and intrinsic interest in an activity, as defined primarily by the children’s selection of worksheets during choice periods. Unlike previous research in this area, it was deemed of critical importance to employ experimental conditions closely analogous to token economy programs. Thus, the subjects in this study were children in a special class for remediation in mathematics. The treatment condition lasted several weeks, rather than a few days. The rewarded response was accuracy in completing math computation worksheets, a target behavior deemed educationally relevant for these children. Moreover, collection of baseline data before the introduction of the treatment package permitted an analysis of whether the reward system actually had a reinforcement effect. Under such conditions, what is the effect of rewarding children for accurate computation during one part of the school day on their selection, accuracy, and enjoyment of the same types of worksheets later in the day when no rewards were available?
Subjects

Six children, four boys and two girls, served as subjects. They ranged in age from 9 years 5 months to 13 years, 2 months. The average age was 11 years. Two of the children were from the third grade, two from the fourth grade, one from the fifth, and one from the sixth grade. The children were identified by their mathematics resource teacher as needing added assistance in arithmetic computation. On the Key Math Diagnostic Test (Connolly, Nachtman, & Pritchett, 1971), they scored an average of 1.4 grade levels below average. These six subjects were selected from a larger group of students needing special assistance on the basis of scheduling compatibility with the experimenter, and the agreement of their parents to have their children participate in the program.

Personnel

The experimenter who served as tutor in this project was a 26 year-old male doctoral candidate in Educational Psychology, having previous experience in applied behavior analysis as a classroom teacher with special needs children. Two language arts tutors
also assisted in data collection.

**Setting**

Data were collected in two different settings. The actual tutoring sessions were conducted in the math resource room of the school. The 6 children were tutored individually at a table removed from their regular resource room area in order to limit distractions. When being tutored, the children were seated so that they were unable to see the rest of the room.

Data on the children's delayed choice of worksheets were collected at a later time in the day, during an individual tutoring session with the language arts teacher. The setting for the collection of these data varied with the individual child. For four of the children, these tutoring sessions occurred in the language arts resource room. For the other two children, the choice of worksheets occurred in a tutoring room next to the regular classroom.

**Materials**

Each child's specific mathematics skill level was pinpointed by means of the Key Math diagnostic profile, as well as in consultation with the math resource teacher. Based on that assessment, individually tailored computation worksheets were designed
in the areas of addition, subtraction, multiplication, division, and simple word problems. Each addition, subtraction, multiplication, or division worksheet contained 25 problems. The word problem worksheets contained only 15 problems, due to the fact that such problems took a longer period of time to complete.

To enhance the discriminability of the problem category, each child worked on three differently colored worksheets, each correlated with specifically different problem categories. Each of the worksheets was placed in a folder of the same color as the worksheet, as noted in Table 2.

Immediately following the completion of a worksheet, each of the children rated their enjoyment of the task on a 5-point scale. To assist the children in this decision, they were shown a sheet of paper upon which were drawn a row of five faces in varying degrees of smiles and frowns (see Figure 1). The child indicated his enjoyment of the task by choosing a number that corresponded to one of the five faces.

Other special materials used in this study included a 10-cm dial stopclock with large distinct numerals and gradations, a Sony tape recorder for taping all of the tutoring sessions, and a three minute sand-in-glass egg timer.
<table>
<thead>
<tr>
<th>Multiplication and Division</th>
<th>Addition and Subtraction</th>
<th>Word Problems</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition and Subtraction</td>
<td>Word Problems</td>
<td>Multiplication and Division</td>
<td>5</td>
</tr>
<tr>
<td>Multiplication</td>
<td>Addition and Subtraction</td>
<td>Word Problems</td>
<td>4</td>
</tr>
<tr>
<td>Word Problems</td>
<td>Multiplication and Division</td>
<td>Addition and Subtraction</td>
<td>3</td>
</tr>
<tr>
<td>Word Problems</td>
<td>Addition and Subtraction</td>
<td>Multiplication</td>
<td>2</td>
</tr>
<tr>
<td>Addition</td>
<td>Word Problems</td>
<td>Subtraction</td>
<td>1</td>
</tr>
<tr>
<td>Pink Worksheet</td>
<td>Blue Worksheet</td>
<td>Yellow Worksheet</td>
<td>Subject</td>
</tr>
</tbody>
</table>

Table 2: Types of Problems on Each Worksheet
Figure 1. Rating chart used by subjects after completing each worksheet.
General procedures

Each of the six children was seen individually in the mathematics resource room four to five days per week. During that time, each child worked on one worksheet for three minutes from each of the three different colored folders. For each type of worksheet, the difficulty of the problems paralleled the children's regular math program. Each child progressed from one level of difficulty to another when 80% of the problems on the worksheet were attempted and 80% accuracy was achieved.

During the tutoring session, each of the three worksheets was presented in random sequence, and the child was instructed to complete accurately as many problems as possible in three minutes. The tutor then started the time-clock, which the child could see at all times. The tutor did not interact with the child at any time during the three minute interval when he was working on the problems. Rather, the tutor was seated next to but slightly behind the child, recording on-task behavior. When the three minute limit was over, the tutor then either gave points for each problem solved correctly (Token Reward condition worksheet), positive verbal feedback (Verbal Reward condition worksheet), or simply said
'OK' or 'Un-hun' (No Reward condition worksheet). After completing each worksheet, the child rated enjoyment of the task on a 5-point scale. The worksheet was then placed back into the appropriately colored folder, and the same procedure was followed for the remaining worksheets.

After the third worksheet was completed, the folders were again taken out, and placed in front of the child. The child was instructed to choose one of the three worksheets to complete back at his own desk. The egg-timer was presented as a reminder of the three minute time limit. When this fourth worksheet (Immediate Choice worksheet) was completed, the child simply put the sheet back into the folder without any interaction with the experimenter.

At a later time in the day, the child had a second opportunity to choose one of the three types of worksheets (Delayed Choice worksheet). This choice was offered by the language arts tutor who followed the same procedure used by the math tutor. Again, the language arts tutor did not interact with the child during work on the problems. When the three minute limit was over, the sheet was returned to the folder without any comment (other than 'ok' or 'un-hun') by the tutor. Thus, for neither the
immediate nor delayed choice worksheets did the child receive points or positive verbal feedback. The tutor who offered this delayed choice was not aware of which worksheets were in each experimental condition.

Design

A multielement experimental design (Ulman & Sulzer-Azaroff, 1975) was utilized to assess the relationship between various reward systems, and the children's performance and interest in each of the three types of worksheets both during the tutoring session, and during immediate and delayed choice periods. During the tutoring session, each of the three types of worksheets was associated with one of three conditions—no reward, verbal reward, or token reward. However, during the immediate and delayed choice periods, all worksheets were completed under the no reward condition. Thus, this design permitted an analysis of the effects of rewards given during one part of the school day on the children's interest (rating and choice) in the math activity at a later time, when no special rewards were forthcoming.

For worksheets completed in the no reward condition, the child was simply asked to rate his
worksheet immediately upon the end of the three minute interval. The tutor neither corrected any of the problems, nor commented in any way (other than "ok" or "un-hun") on any aspect of the work.

For worksheets completed in the verbal reward condition, the tutor delivered between two and four positive praise comments as he corrected the child's worksheet. Such comments were both specific (e.g., "you're remembering to carry the 1") and general (e.g., "you're really doing well on these addition problems") praise statements.

For worksheets completed in the token reward condition, children earned one point for each problem solved correctly. The problems were corrected immediately after the child completed the worksheet. Each day's points were recorded on a graph that was then placed inside the child's folder. At the end of the week, the children had the opportunity to spend their points at a special store set up in the resource room. In general, the children redeemed their points for such items as baseball cards, balloons, marbles, toy soldiers, small dolls, or soap bubbles. The children were allowed to save their points from one week to the next, although few chose to do so.
On any given day, the number of praise comments delivered during the verbal reward condition was yoked to the number of praise comments emitted in the token condition. For example, if three positive comments were made during the verbal reward condition, the same number of comments was made in the token condition. On days when the token reward condition preceded the verbal reward condition, the number of praise comments delivered during the former was simply yoked with the latter.

**Experimental phases**

**Baseline phase.** During the baseline phase, all three types of worksheets were completed under the no reward condition. The purpose of this phase was to assess the children's interest and performance in the activities independent of any specific verbal or token rewards. After the child completed the three worksheets during the tutoring session, a fourth was chosen to complete alone (immediate choice worksheet). At a later time in the day, the fifth worksheet was chosen in the presence of another tutor (delayed choice worksheet). Thus, the child's completion of all five worksheets was followed by neither positive verbal feedback, nor any point system. This phase terminated on the same day for all of the children,
<table>
<thead>
<tr>
<th>Choice Periods</th>
<th>Immediate Choice Activity</th>
<th>Medium Choice Activity</th>
<th>Low Choice Activity</th>
<th>Delayed Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
</tr>
<tr>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
</tr>
<tr>
<td>Token Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
</tr>
<tr>
<td>Verbal Reward</td>
<td>Verbal Reward</td>
<td>Verbal Reward</td>
<td>Verbal Reward</td>
<td>Verbal Reward</td>
</tr>
<tr>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
</tr>
<tr>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
<td>No Reward</td>
</tr>
</tbody>
</table>

**Table 3**: Reward conditions associated with each experimental phase.
after five to ten tutoring sessions (average of 7.5) had been completed. The variability in the number of baseline sessions was due to the fact that it was not possible to tutor each child every day.

**Token-Low Choice phase.** The worksheet that each child had chosen least frequently in the baseline phase was designated the low choice worksheet for the rest of the experiment, and placed in the token reward condition for this phase. That is, the children would now earn points for each problem solved correctly on the worksheet chosen least frequently in the baseline phase. The purpose of this phase was to assess the effect of the point system (operative only during the tutoring sessions) not only on that worksheet, but also on the immediate and delayed choice of worksheets. In cases where there was a tie in terms of least chosen activity, the worksheet that the child had rated lower on the 5-point scale was assigned to this condition.

The worksheet that each child had chosen second most frequently during the baseline phase was designated the medium choice worksheet for the rest of the experiment, and placed in the verbal reward condition for this phase. That is, between two and four positive statements were delivered by the tutor as he corrected
the worksheet in this condition. For all of the subjects except subject 3, this condition was introduced a few days before the token reward condition. The purpose of this staggered introduction of reward systems was to ascertain the effect of the verbal reward condition without changing the other two conditions. Since subject 3 was absent for the few days when the condition was introduced with the other children, she was introduced to verbal and token conditions simultaneously.

The worksheet that each child had chosen most frequently in the baseline phase was designated the high choice worksheet for the rest of the experiment, and placed in the no reward condition for this phase. Since the students had already been completing worksheets during Baseline in the no reward condition, there was no change in consequences for this activity.

In summary, during the Token-Low Choice phase, the children received points for accurately solving problems that they had previously chosen least frequently, positive verbal feedback for completing worksheets that they had previously chosen with a medium level of frequency, and no points or verbal feedback for completing worksheets that they had chosen most frequently during Baseline. However, during this phase,
as well as during all phases, the children's immediate and delayed choice of worksheets continued to be completed under the no reward condition. For example, while a child might earn points for each correct answer on the addition worksheet during the tutoring session, points could not be earned if that same activity was selected during either of the choice periods.

For subjects one, two, and three, this was the last phase of the experiment, and lasted 14, 14, and 17 days respectively. The termination of this phase corresponded with the end of the school year. For subjects four, five, and six, this phase lasted 11, 13, and 12 days respectively.

**Token-High Choice phase.** The purpose of this phase was to ascertain the consequences of switching the token reward condition from the least favored to the most favored activity. Thus, subjects four, five, and six, who during the Token-Low Choice phase had demonstrated the highest preference for a particular activity, were placed in this phase. During the Token-High Choice phase, the high choice worksheet was switched from the no reward to the token reward condition. The medium choice worksheet continued to be completed under the verbal reward condition.
As in the two previous phases, the children's completion of both immediate and delayed choice worksheets continued to be completed under the no reward condition.

**Behaviors measured**

For each of the six children, the following behaviors were recorded daily.

1. **Immediate choice.** After completing the three worksheets during the regular tutoring session, each child was instructed to choose a fourth worksheet to complete in the resource room. The three types of worksheets, similar in difficulty to the just completed tasks, were placed in front of the child. Choosing one worksheet, the child then returned to his desk with the three minute egg-timer. When the three minutes were up, the child returned the sheet and rated enjoyment of the task by placing a number in the upper left hand corner of the sheet.

2. **Delayed choice.** At a later time in the day, the children had a second opportunity to choose one of the three types of worksheets. This choice was offered to them by the language arts tutor. The conditions under which they completed this delayed choice worksheet were identical with that of the immediate choice. They had three minutes to work
on the worksheet, and rated their enjoyment of the activity on the same 5-point scale.

The child's preference for worksheets in the absence of rewards was intended to serve as a measure of the effect of rewards given during one part of the day upon the children's behavior at other times of the day. The two choice periods, immediate and delayed, were implemented to assess the stability of such preferences.

3. **Subjective rating of worksheets.** Immediately after completing a worksheet in both the tutoring and choice periods, each of the children rated their enjoyment of the task on a 5-point scale as described in the materials' section. The rationale for measuring this behavior was to ascertain the relation between the children's verbal reports of their enjoyment of the task, and the worksheet they actually selected during choice times.

4. **On-task behavior during tutoring sessions.** The children's on-task behavior was measured while they were working on each of the three worksheets during the tutoring session. On-task behavior was defined as any time the child was looking at his worksheet, or counting out loud. A momentary time sample procedure (Powell, Martindale, & Kulp, 1975) was utilized to assess this behavior. The three
minute work session for each worksheet was divided in 36 5-second time blocks. A stopclock with distinct numerals and a 10 cm dial indicating seconds was placed on a small shelf immediately above the desk. At the end of each 5 second interval, the tutor recorded whether the child was on-task at that moment by placing a mark on the scoring sheet. The tutor was sitting to the left and slightly behind the child. The child was thus unable to see the tutor's recording. However, before initiating the system, the procedure was explained and demonstrated to each child.

5. **Number of math problems attempted.** The number of problems attempted for each of the worksheets both during the tutoring and choice sessions was recorded for each child. A problem was recorded as attempted if the child placed any number in the vicinity of the problem.

6. **Percentage of problems solved correctly.** The percentage of problems solved correctly was attained by dividing the number of problems solved correctly by the number of problems attempted. This figure was multiplied by 100 to yield a percentage of problems solved correctly.
Reliability

An observer naive to the experimental manipulations observed each child at least once during each experimental phase, and simultaneously recorded on-task behavior. The observer was seated several feet behind and to the side of each child, seated in such a way that it was not possible to note the experimenter's recordings. It was considered important that the observer not be aware of which worksheets were correlated with each reward condition. For that reason, the observer left the area immediately after the child completed a worksheet, but before the tutor dispensed either verbal or token rewards. Reliability was calculated by dividing the number of agreements by the total agreements plus disagreements, then multiplying the fraction by 100. An interval by interval comparison was conducted, with a disagreement being any scored interval in which the observer and experimenter differed in their recording. An interval was scored as an agreement if both observer and experimenter recorded the scored interval the same way.

The overall reliability of on-task recordings for the six subjects was 91%. The reliability for subject one was 87% (range of 78% to 100%), subject
two was 91% (range of 85% to 100%), subject three was 96% (range of 92% to 100%), subject four was 91% (range of 83% to 97%), subject five was 39% (range of 36% to 94%), and subject six was 93% (range of 87% to 96%).

A random sample of 15 worksheets from each of the six subjects was also scored by an observer for the number of problems attempted. Reliability was calculated by dividing the number of agreements by the number of agreements plus disagreements, and then multiplying the fraction by 100. An agreement was defined as any math problem in which both the rater and the experimenter recorded it as attempted. The overall reliability of math problems attempted for the six subjects was 97.6%, with a range of 92% to 100%. The same sample of 90 worksheets was also scored for the number of problems solved accurately, with reliability calculated on a problem by problem basis using the above method of calculation. The overall reliability of math problems solved accurately for the six subjects was 98.4%, with a range of 91% to 100%.
Results

Choice

The choice, rating, and on-task behavior of each child is summarized in Table 4. The combined average choice reflects the number of choices made in both the immediate and delayed periods, divided by the total number of choices.

Points delivered contingently during the tutoring session resulted in an increase in the selection of the low choice activity in the Token-Low Choice phase for each of the six subjects. Whereas, as a group, the low choice activity was selected 14% of the time during Baseline, it was chosen 29% of the time during the Token-Low Choice phase. For the three subjects in the Token-High Choice phase, the cessation of the token system resulted in a decrease in the choice of the worksheet associated with that condition, dropping an average from 26% to 5%, 5% lower than Baseline levels. Figures 2 through 6 reflect the trend of the choices for each of the six subjects.
<table>
<thead>
<tr>
<th></th>
<th>%85</th>
<th>%93</th>
<th>%96</th>
<th>%88</th>
<th>%87</th>
<th>%71</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average choice</td>
<td>4.0</td>
<td>4.7</td>
<td>4.0</td>
<td>4.0</td>
<td>4.4</td>
<td>4.0</td>
</tr>
<tr>
<td>Delayed choice</td>
<td>4.0</td>
<td>4.0</td>
<td>3.7</td>
<td>3.3</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Immediate choice</td>
<td>4.0</td>
<td>4.0</td>
<td>4.5</td>
<td>5.0</td>
<td>4.5</td>
<td>4.0</td>
</tr>
<tr>
<td>Percentage on-task</td>
<td>56%</td>
<td>60%</td>
<td>60%</td>
<td>55%</td>
<td>60%</td>
<td>54%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>%96</th>
<th>%96</th>
<th>%96</th>
<th>%96</th>
<th>%96</th>
<th>%96</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average choice</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Delayed choice</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Immediate choice</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Percentage on-task</td>
<td>60%</td>
<td>60%</td>
<td>60%</td>
<td>60%</td>
<td>60%</td>
<td>60%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>%54</th>
<th>%59</th>
<th>%55</th>
<th>%58</th>
<th>%62</th>
<th>%65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average choice</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Delayed choice</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Immediate choice</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Percentage on-task</td>
<td>37%</td>
<td>37%</td>
<td>37%</td>
<td>37%</td>
<td>37%</td>
<td>37%</td>
</tr>
</tbody>
</table>

Table 4
Choice, rating, and on-task behavior of subjects 1-6
<table>
<thead>
<tr>
<th>Subject</th>
<th>Choice</th>
<th>Verbal Token</th>
<th>Token-Low Choice</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 4 (cont.)
<table>
<thead>
<tr>
<th>%</th>
<th>94%</th>
<th>70%</th>
<th>65%</th>
<th>60%</th>
<th>57%</th>
<th>55%</th>
<th>53%</th>
<th>50%</th>
<th>47%</th>
<th>45%</th>
<th>43%</th>
<th>40%</th>
<th>37%</th>
<th>35%</th>
<th>33%</th>
<th>30%</th>
<th>27%</th>
<th>25%</th>
<th>20%</th>
<th>16%</th>
<th>10%</th>
<th>6%</th>
<th>2%</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Percentage on-Task | 4-G  | 4-6  | 4-7  | 4-8  | 4-9  | 4-10 | 4-11 | 4-12 | 4-13 | 4-14 | 4-15 | 4-16 | 4-17 | 4-18 | 4-19 | 4-20 | 4-21 | 4-22 | 4-23 | 4-24 | 4-25 | 4-26 |
|---------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
|                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |

*Table 4 (cont.)*
Figure 2. Subject 1: Percent of on-task behavior during tutoring. Worksheets selected during immediate (●) and delayed (○) choice periods.
Figure 3. Subject 2: Percent of on-task behavior during tutoring. Worksheets selected during immediate (●) and delayed (○) choice periods.
Figure 4. Subject 3: Percent of on-task behavior during tutoring. Worksheets selected during immediate (●) and delayed (○) choice periods.
Figure 5. Subject 4: Percent of on-task behavior during tutoring. Worksheets selected during immediate (⬤) and delayed (○) choice periods.
Figure 6. Subject 5: Percent of on-task behavior during tutoring. Worksheets selected during immediate (○) and delayed (●) choice periods.
Figure 7. Subject 6: Percent of on-task behavior during tutoring. Worksheets selected during immediate (●) and delayed (○) choice periods.
For five of the six subjects, the verbal reward condition did not increase the selection of that activity during the Token-Low Choice phase. Whereas, as a group, the medium interest activity was chosen 20% during Baseline, it was selected 18% when it was associated with positive verbal feedback during the tutoring session. During the Token-High Choice phase, when the medium choice interest continued to be associated with the verbal reward condition, the choice of this worksheet further decreased from 4% to 2%.

For five of the six subjects, the selection of the high choice activity decreased from Baseline to the Token-Low Choice phase, dropping an average from 66% to 53%. However, for the three subjects in the Token-High Choice group, the selection of the high choice worksheet increased to 93% when it was associated with the token reward condition during the tutoring session.

**Ratings—Tutoring Session**

While the low choice worksheet was rated only 3.5 during the no reward condition, its average rating increased to 4.5 when the same worksheet was associated with the token reward. When the low choice activity was returned again to the no reward
condition, its average rating decreased to 4.1. While the high choice activity received an average rating of 4.2 during Baseline, and a rating of 4.1 during the Token-Low Choice phase, its rating increased to 4.7 when it was associated with the token reward in the third phase.

Ratings—Choice Periods

The low choice activity, selected only 14% of the time during Baseline, also received the lowest rating in comparison with other worksheets when it was chosen—an average of 3.5 on the 5-point scale. However, when it was associated with the token reward during the tutoring session (Token-Low Choice phase), its rating during choice periods increased from 3.5 to 4.6. When the low choice activity was again associated with the no reward during the tutoring session, its average decreased to 4.5.

The average rating of the high choice activity was 4.6 during Baseline, and 4.1 during the Token-Low Choice phase. However, during the Token-High Choice phase, when the high choice activity was associated with the token reward, its average rating increased to 4.8
On-Task Behavior

While the students were on-task 74% of the time with the low choice activity during the no reward condition, that percentage increased to 96% when it was associated with the token reward condition. When the same low choice worksheet was returned to the no reward condition during the Token-High Choice phase, on-task behavior decreased an average from 95% to 65%. Correspondingly, on-task behavior increased from 87% to 94% for the high choice worksheet during the Token-High Choice phase.

Problems Attempted/Accuracy

The average number of problems attempted and solved accurately for each of the six subjects during tutoring and choice periods is reported in Table 5. For the low choice worksheet, accuracy during the no reward condition was 74%, but increased to 87% under the token reward condition during tutoring periods. During the Token-High Choice phase, accuracy decreased from 85% to 70% for the three students in that condition. During both the Token-Low Choice and Token-High Choice phases, the greatest number of problems attempted were worksheets associated with the token reward condition.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
<th>Attempt</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Problems attempted and accuracy during tutoring and choice periods.
In summary, low choice worksheets completed during the token reward condition were chosen more frequently, rated more highly during tutoring and choice sessions, solved more accurately, and attended to a greater percentage of time than the same worksheets completed under the no reward condition in the Baseline phase.
Discussion

Extrinsic rewards (i.e., point system) given during the tutoring periods increased both the selection and rating of low and high choice worksheets later on in the day when no rewards were available. When the low choice worksheets were associated with the token reward condition during tutoring, all six subjects selected that worksheet more frequently during choice periods, even though no rewards were ever given during the choice periods. Students also rated that worksheet more highly, and performed more accurately than they did during the Baseline phase.

The token reward had a clear reinforcement effect on the children's performance on the low choice activity during tutoring. The six children were on-task more frequently (an increase from 74% to 96%), attempted more problems (an increase from 14.7 to 18.5), and performed more accurately (an increase from 74% to 87%) than they did during the Baseline phase. They also rated the worksheets as more enjoyable (an increase from 3.5 to 4.5). Thus, the reinforcement of an activity during one part of the day served to increase rather than decrease
interest in that activity when rewards were unavailable.

When the token reward condition was associated with the high choice worksheet for three of the children, an analogous effect was noted. Again, a reinforcement effect was documented during the tutoring session---on-task behavior increased from 87% to 94%, accuracy increased from 82% to 88%, and the rating of the activity increased from 4.3 to 4.7. However, the number of problems attempted decreased from 16.2 to 15.6, perhaps due to the fact that the problems were getting more difficult as the children progressed. When the high choice activity was associated with the token reward, the worksheet was chosen more frequently and rated more highly during the choice periods, even though no rewards were given during those choice periods.

Thus, for both low and high choice worksheets, extrinsic rewards in the form of a token system served to increase intrinsic interest in the task, as measured by the students' selection of activities, and rating of enjoyment of that activity when no rewards were offered.

The effect of the verbal reward is more ambiguous. First, it should be noted that the verbal reward was not a reinforcer for five of the six children---i.e.,
no clear reinforcement effect was noted. Examination of several dependent measures supports this conclusion. The percentage of on-task behavior during the tutoring session for the medium choice worksheet was 83% during Baseline, but dropped to 81% during the Token-Low Choice phase. This downward trend in on-task behavior was noted for five of the six children. Four of the students also attempted fewer problems during the Verbal reward condition, but four of the children did solve the problems more accurately during the tutoring session. Five of the six children rated the worksheets in the verbal reward condition lower than during Baseline. Although there are some conflicting trends, it appears as if the verbal reward condition did not have a reinforcing effect.

This lack of reinforcement effect during the tutoring session may account for the fact that five of the six students chose the medium choice worksheet less frequently during the Token-Low Choice phase. During the Token-High Choice phase, when a verbal reward continued to be associated with the medium choice worksheet, it was only selected 2% of the time.

What effect did the no reward condition have on interest in the activity? During both Baseline
and the Token-Low Choice phases, the high choice worksheet was associated with the no reward condition. As evidenced by the number of problems attempted, accuracy, and the percentage of on-task behavior, there was a decrease in interest in the task—although some individual differences between subjects and between measures are noted. The selection of worksheets during the choice periods also reflects this decreased interest. Five of the six children selected the high choice worksheet less frequently during the Token-Low Choice phase than they did during Baseline. During the Token-High Choice phase, the no reward condition was associated with the low choice worksheet for three of the subjects. This association resulted in a decrease in on-task behavior, decrease in the number of problems attempted (subjects 4 and 5), decrease in accuracy (subjects 4 and 6), and a decrease in rating (subjects 4 and 6). During the choice periods, the association of the no reward condition with the low choice worksheet resulted in a drop in its selection for all three children, from 26% to 5%.
Possible explanation of effects

These data are, in general, in marked contrast to previous results in this area. Rather than decreasing interest in the task, the point system actually increased interest in the activity. What factors may be responsible for this increased interest?

Within an operant framework, the above results can be described in terms of stimulus and response generalization. Stimulus generalization refers to a situation where "a response conditioned in the presence of one stimulus also occurs in the presence of other, physically different although related stimuli" (Nevin, 1973, p. 115). Within the context of the present study, the children's generalization of accurate responding from the tutoring session (where accuracy was rewarded) to the choice periods (where accuracy was not rewarded) represents an example of such stimulus generalization. Analogously, evidence of response generalization—concurrent changes in responses other than in the target (i.e., rewarded) behavior—can also be found. While the number of problems solved correctly was rewarded during the tutoring session, systematic response changes were also noted in other nontarget behaviors—
e.g., percentage of on-task behavior, self-report ratings, and most importantly for the purposes of the present study, selection of worksheets during the choice periods.

Such response generalization has been reported with increasing frequency in token economy programs (Kazdin, 1975). Twardosz and Sajwaj (1972) reported that rewarding in-seat behavior also increased such nontarget behaviors as play skills and social interaction. Kubany, Weiss, and Sloggett (1971) reported that eliminating the disruptive classroom behavior of a 6 year-old also was correlated with increases in such behaviors as punctuality in coming to class after recess. However, the growing number of reports of such response generalization, perhaps due to a greater proclivity of researchers to measure collateral responses other than the target behavior, should not be interpreted as evidence that such generalization is a predictable concomitant to token economies.

A description of the children's choice behavior as an example of response generalization should not be confused with an explanation of that behavior. The term "generalization" refers to a description of an empirical relationship, and does not necessarily constitute an explanation of that phenomenon. To
qualify as an explanation, it is necessary to identify the critical features of the environmental situation that are functionally related to the specific responses—in this case, selection of worksheets during the choice periods.

What are the necessary and sufficient environmental conditions responsible for response generalization in the present study? Two alternative research approaches might be pursued in investigating that question.

One approach is to focus attention on determining what specific experimental conditions are related to what types of response generalization. This strategy might assess the import of such factors as the magnitude of the reward, nature of the target behavior, type of reward, schedule of reinforcement, nature of the subject population, and contingencies operating in the natural (i.e., unprogrammed) environment. Over time, this approach may serve to identify the conditions under which specific behaviors are members of the same response class—i.e., behaviors that functionally correlate with each other such that changes in one response may be predicted from changes in another response. This inductive approach essentially concerns itself with determining the
functional relation between environmental stimuli and particular responses.

Within that research approach, an explanation of the results of the current study would be given in terms of a specification of the relation between independent and dependent variables. That is, if a token reinforcer is delivered contingent upon accuracy of completion of math problems, then selection of the rewarded worksheet will increase at later times when rewards are unavailable. Asking why that effect occurs is synonymous with requesting more information on the necessary and sufficient conditions that are correlated with that effect. That is, would the same relation between reinforcer and selection of worksheets hold true if the reward was larger, smaller, different, or delivered on an intermittent schedule? Would the same effect occur with children who were older, younger, retarded, or gifted?

A second research orientation works within some theoretical framework (e.g., attribution theory) to focus attention on the relation between some environmental event and a hypothetical construct. This deductive strategy differs from the first approach in that there is a greater willingness to employ hypothetical constructs as explanatory aids. However,
such distinctions are more a matter of emphasis, rather than qualitative differences. All scientific approaches are, to some degree, both deductive and inductive in that a researcher is always working within some theoretical framework, although it may be implicit. Correspondingly, inferences are an inevitable aspect of any scientific methodology. The question is not whether inferences are made, but rather under what conditions they are made.

Attempts at explaining the results of the current study within this second orientation may be facilitated by comparisons with other research employing a multiple trial procedure. Reiss et al. (1975) also found that an extrinsic reward did not decrease intrinsic interest, although several significant methodological limitations have been noted with that study (Lepper & Greene, 1976). Only Greene et al. (1976), using a multiple trial procedure with high and low interest tasks, found a decrease in interest as a function of the reward contingencies. In attempting to explain the apparent discrepancy between that study and the current results, several differences could be noted—e.g., experimental design, nature of the token program, subject selection, etc. However, one suggested difference of major importance is the nature of the
rewarded behavior. Greene et al. (1976) rewarded the amount of time that children spent on a particular math activity and not performance on that activity. A reinforcement effect was demonstrated in that the amount of time spent on certain target activities increased with the introduction of the reward. However, statistical analyses did not reveal any differences between rewarded and non-rewarded groups in either rate of completion or accuracy of math work. In contrast, the current investigation involved the reward of rate and accuracy (number of problems correct within a three minute limit). Within self-perception theory, the nature of the rewarded task may be quite critical. As noted earlier, it was suggested that rewards have a dual function—controlling (getting someone to do something he doesn't want to do) and feedback (communicating the activity is valued). Rewarding accuracy of performance, as was done in the current study, and many token economy programs, would seem more likely to increase a child's sense of competence and mastery, thus having more of a feedback function. In contrast, simply rewarding the amount of time spent on an activity may be perceived as more controlling or manipulative. Irrespective of this hypothesized difference, the
dependent measure used by Greene et al. (1976) is not recommended in most token economies, thus limiting the external validity of their results. Thus, in accounting for the results of the current study, it is suggested that the nature of the target behavior may be a critical factor in facilitating response generalization—namely, that accuracy of performance and choice of worksheets are members of the same response class.

**Contribution of the present study**

The practical import of the current investigation is to suggest that the generalization that "token economies may lead to token learning" (Levine & Pasnacht, 1974, p. 816) is an inappropriate extrapolation from the available research. Token economy programs have been successfully implemented with a wide variety of subjects, settings, and behaviors (e.g., see Kazdin & Bootzin, 1972; O'Leary & Drabman, 1971). There is nothing in the available evidence to suggest that rewards inherently decrease 'intrinsic' interest. While problems of implementation, maintenance, and generalization remain quite real, it is critical that treatment packages of proven efficacy not be inadvertently discarded.
However, there are several important limitations to the current investigation that should be noted. The task involved a relatively brief academic assignment, and it remains unclear whether intrinsic interest measured later on in the day would have been affected had the response requirements for rewards been higher (i.e., longer tutoring session). The token system was also in effect for a brief period of time, prohibiting an analysis of whether interest in the task was affected by an extended duration of the token system. It also might be argued that a measure of prime importance is the interest in the task once the reward system was removed. The abrupt removal of the token system for subjects four, five and six does provide some limited data on that issue. For two of the three subjects, the percentage of on-task behavior decreased below Baseline levels. However, such critical questions as the permanence of this response suppression, as well as the effect of the gradual removal of the token system remains unanswered by this study.

But the maintenance of behavior change in token economies remains a troublesome problem. In his review of token economy research, Kazdin (1975) has noted that "to many, the long-term maintenance of
target behaviors developed through token reinforcement represents an unfulfilled promissory note. The highest priority, in the opinion of the author, is for empirical demonstrations of effective maintenance strategies (p. 263-264)." A self-perception analysis of many proposed techniques for fostering behavior maintenance and generalization (e.g., self-control procedures, self-instruction training, scheduling intermittent reinforcement) may yet prove helpful in developing a behavioral technology of response maintenance.
References


Connolly, A. J., Nachtman, W., & Pritchett, E. M.  


Deci, E. L. The effects of contingent and non-contingent rewards and controls on intrinsic motivation. Organizational Behavior and Human Performance, 1972, 8, 217-229. (b)


Greene, D., & Lepper, M. R. *How to turn play into work.* Psychology Today, 1974, 8, 49-54.


Harlow, H. F. *Motivation as a factor in the acquisition of new responses.* Nebraska symposium on motivation, 1953, 1, 24-49.


Kubany, E. S., Weiss, L. E., & Sloggett, B. B.  


Twardosz, S., & Sajwaj, T. Multiple effects of a procedure to increase sitting in a hyperactive, retarded boy. *Journal of Applied Behavior Analysis, 1972, 5*, 73-78.


