DOCUMENT RESUME

ED 193 331 TM 800 649

AUTHOR Schilling, Deanna E.
TITLE Reward Experience, Socioeconomic Status, and Sex: Exploring Parameters of the Overjustification Effect.
PUB DATE Apr 79
NOTE 35p.: Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Grade 1: Hypothesis Testing: Motivation Techniques: Primary Education: *Reinforcement: *Rewards: *Sex Differences: *Socieceonomic Status: Student Motivation
IDENTIFIERS *Overjustification

ABSTRACT

The overjustification hypothesis predicts decreased intrinsic motivation when persons are paid to perform an interesting task. The factors of reward experience, socioeconomic status (SES), and sex are examined while testing conflicting predictions of the hypothesis and reinforcement theory. Children from grade 1 at two public elementary schools worked on a counting task. Half of the children attended a school which had a 100% minority enrollment (low SES), and half attended a school which had a 95% Caucasian enrollment (middle SES). A Baseline 1, treatment session, Baseline 2 paradigm was used to assess the effects of both reward introduction and withdrawal. Reward experience (group), SES, and sex were manipulated in a 3 x 2 x 2 design. Subjects who earned a reward during the treatment session for maintaining Baseline 1 output levels, significantly outperformed never-rewarded controls during Baseline 2. Control subjects initially highest on motivation measures declined significantly from Baseline 1 to Baseline 2: comparable subjects in the reward groups showed no change. In addition, reward-group subjects that were initially lowest in motivation significantly increased their output and time on task across sessions. (Author/RI)

Reward Experience, Socioeconomic Status, and Sex: Exploring Parameters of the Overjustification Effect

Deanna E. Schilling

University of California at Davis

This article is based on a doctoral dissertation submitted to the Department of Psychology, University of California at Davis.

A presentation based on this research was made at the 1979 American Educational Research Association meeting, Session 16.15 Process-Product Research on Teaching I, Division C

Abstract

The overjustification hypothesis predicts decreased intrinsic motivation when persons are paid to perform an interesting task. This study examined the factors of reward experience, socioeconomic status, and sex while testing conflicting predictions of the hypothesis and reinforcement theory. Children worked on a counting task in a Baseline 1, treatment session, Baseline 2 design. Subjects who earned a reward during the treatment session for maintaining Baseline 1 output levels, significantly outperformed never-rewarded controls during Baseline 2. Control subjects initially highest on motivation measures declined significantly from Baseline 1 to Baseline 2; comparable subjects in the reward groups showed no change. In addition, reward-group subjects that were initially lowest in motivation significantly increased their output and time on task across sessions.

Reward Experience, Socioeconomic Status, and Sex: Exploring Parameters of the Overjustification Effect

There are four combinations possible when high and low levels of extrinsic reward are considered jointly with high and low levels of intrinsic reward. Of these, the bulk of recent attention has been directed toward the combination of high extrinsic/high intrinsic reward that occurs when an individual is paid for performing an interesting task. Self-perception theory (Bem, 1965, 1967, 1972) and personal causation theory (de Charms, 1968) maintain that rewarding an already-interested individual may cause a decrease in intrinsic motivation, an outcome sometimes referred to as an "overjustification effect" (cf. Bem, 1972). Such an effect is believed to occur either because rewards place the receivers in a dependent position causing them to feel less free (de Charms), or because individuals infer they did not want to do an activity when external reinforcement contingences are particularly salient (Bem, 1972, p. 39). In contrast, the more traditional "additive" view of motivation (cf. Notz, 1975) assumes that intrinsic and extrinsic motivation summate; thus, extrinsically rewarding a person for doing an interesting task should lead to the greatest possible motivation (cf. Arnold, 1976; Davis, Settlage, & Harlow, 1950; Hedges, 1972).

Relatively few studies[1] have yielded results which clearly support the overjustification or additive positions. Instead, rewards have more often been found to have both enhancing and detrimental effects on measures of intrinsic motivation as a function of such factors as type of reward

(Anderson, Manoogian, & Reznick, 1976); type of task (Calder & Staw, 1975); time of payment (Deci, 1972); sex (Deci, Cascio, & Krusell, 1975); whether money was a "natural consequence" of the task (Kruglanski, Riter, Amitai, Margolin, Shabtai, & Zaksh, 1975); initial level of interest in the task and expectation of reward (Lepper, Greene, & Nisbett, 1973).

Despite the mixed nature of these results, an increasing number of articles are appearing in which it is warned that the use of rewards may have detrimental effects on motivation in educational settings. It is important to note that the authors of these articles do not always distinguish between high- and low-interest activities (cf. Kruglanski, Alon, & Lewis, 1972; Sorensen & Maehr, 1976). Furthermore, a relatively recent statement of the overjustification hypothesis asserts that it applies to "activities of at least some initial interest to a subject" (Greene, Sternberg, & Lepper, 1976) which is a considerable modification of earlier formulations which stressed the high attractiveness of the activities used when the overjustification effect was obtained. This relaxing of the distinction between high- and low-interest tasks carries with it the implication that rewards have detrimental effects regardless of whether intrinsic motivation is high or low. It is but a small step, then, to recommendations that extrinsic rewards be completely withdrawn from educational settings. Yet, there are reasons to hesitate before making decisions on the basis of past overjustification research, including the rather restricted parameters of these studies. The purpose of the present investigation was to extend previous findings by examining the factors of reward experience, socioeconomic status (SES), and sex, while incorporating certain procedures designed to foster both internal and external validity.

Experimental Design and Hypothesis

Overjustification studies typically give, or promise to give, subjects a reward for merely completing the task. Payment is not tied to any improvement over past performance. (In fact, subjects have been rewarded even when the quality of their performance was significantly lower than that of the nonrewarded subjects, cf. Greene & Lepper, 1974; Lepper, Greene, & Nisbett, 1973.) Yet, in real-life situations, rewards are rarely given for whatever behavior the person chooses to display. There is some basis for thinking a reduction in intrinsic motivation may be less likely to occur when improved performance is required as a condition for reward. There is the traditional view of reinforcement theorists that the positive value of rewards accrues to the responses for which they are given, through a secondary reinforcement process. Thus, subjects rewarded for increasing their baseline output may begin to value such increased performance. This valuing would be expected to manifest itself in higher performance levels, when rewards are withdrawn, relative to the performance of subjects earlier rewarded for simply maintaining their baseline levels or of unrewarded controls.

A Baseline 1, treatment session, Baseline 2 paradigm was used to assess the effects of both reward introduction and withdrawal. Reward experience (Group), SES, and Sex were manipulated in a 3 x 2 x 2 design. The experimental groups were: (a) the reward-same group who were offered a reward during the treatment session for maintaining a level of performance (output) they had achieved voluntarily and without reward during Baseline 1 (a condition of overjustification), (b) the reward-more group who were

offered a reward during the treatment session for improving a level of performance (output) they had achieved voluntarily and without reward during Baseline 1. A control group were neither offered nor given a reward during the course of the study.

The hypothesis of the study was that the Baseline 2 output of the reward-more group would be significantly greater than that of the reward-same and control groups as a result of secondary reinforcement. In contrast, the overjustification hypothesis would predict that the Baseline 2 performance of subjects in the reward-same and reward-more groups would be significantly lower than that of the nonrewarded control group.

With the exception of Pierce (1971), relatively little attention has been paid to the SES factor. The subjects of overjustification research have typically been of middle-class background. Since the effects of reward on student motivation are of particular interest to those concerned about the school performance of low-SES children, the SES factor was included for investigation.

Many researchers have not commented on whether sex differences were found or even considered; a few have used male or female subjects only. The sex factor was therefore examined both for control purposes and because it has been relatively neglected in past experimentation. There were no specific hypotheses concerning the SES and sex factors.

Validity issues. The following steps were taken to promote internal and external validity: (a) Subjects in reward groups were allowed to choose from a variety of prizes, rather than being given identical rewards, (b) Whether subjects were above or below the median on measures

of Baseline 1 performance was used as an index of initial intrinsic motivation (cf., Lepper, Greene, & Nisbett, 1973), instead of assuming subjects were equally interested in the activity, (c) A realistic task, rather than a play activity, was chosen to increase the likelihood that the obtained results would be applicable to natural settings.

Method

Subjects

Subjects were drawn from first-grade classes at two public elementary schools. Half of the subjects (15 boys, 15 girls) attended a school which has 100% minority enrollment, with Blacks constituting about 75% of the student body and the remainder being of Mexican descent. Since 88% of the students enrolled at this school come from families receiving welfare assistance, the subjects at this school comprised the low-SES group. The middle-SES subjects attended a school which serves a neighborhood that is approximately 95% Caucasian. The parents of many of the children are in white-collar or highly skilled occupations, and some teach at a nearby state university.

All subjects were first individually screened to determine whether they could count 10 objects, recognize the numerals 1-10, and identify various sets of less than 10 objects. The tasks were too difficult for three low-SES boys, and they were eliminated from the subject pool. Subjects were then randomly assigned to the three groups so that each group consisted of five low- and five middle-SES boys, and five low- and five middle-SES girls. Within each school, control and experimental subjects were selected from different classrooms, after ascertaining that school policy was to assign children at random to classrooms.

Materials

A practice sheet and three 20-page sets of arithmetic papers were prepared, one set for each of the 3 days subjects were seen. Each page contained 10 problems which required the subject to count a number of objects and then circle the numeral between 1 and 10 which was the correct answer. The sets were in different colors so children would not (mistakenly) think they were doing the same pages each day, and contained equivalent problems. Pages within each of these sets were not stapled but, rather, stacked in sequence. Thus, it was possible for a child to finish 20 pages and begin again on page number one of the same set. Since subjects could not possibly complete the stack of papers, it was believed they would feel freer to work for whatever period of time they wished. Pages were numbered with Roman numerals to help prevent children from stopping on a particular page number because they had stopped at that point on a previous day.

Procedure

Overview. Subjects were seen for 3 days in succession, completing Baseline 1 on the first day, the treatment session on the second day, and Baseline 2 on the last day. If, as a result of absences, a total of more than 2 days elapsed between sessions without the subject being seen, the subject was dropped from the experiment and replaced by another child from the subject pool. The time of day and day of week that subjects were seen were approximately equated for the main factors of Group, SES, and Sex. The following measures were taken during each session: (a) number of problems completed (output), (b) seconds on task, (c) seconds on task divided by number of problems completed (seconds per problem),

(d) percent correct, (e) number of problems skipped.

Baseline 1. Children were individually escorted to the experimental room. Conversation was neutral and kept to a minimum. The subjects were seated at a desk, facing a wall. The stack of arithmetic papers was to their right; a box for completed pages was on the left. After the subjects completed a practice page, they were given the following directions:

> Here is a stack of arithmetic papers. You may do as many papers as you like. Take the top one on the pile. When you have finished it, put it in this box. Then if you want to do another one, take the next one off the stack. Try not to skip problems, and try not to skip pages. Remember, you can work on the counting problems as long as you like. When you are ready to stop, tell me, and we will go back to the room.

The experimenter was seated at a desk behind, and slightly to the right, of the subject. From this vantage, it was possible to observe the subject without being seen. A single-pole (off-on) switch was wired to the cord of an electric clock. Based on a previously prepared list of on- and off-task behaviors, total time on task for subjects was determined by switching the clock on whenever subjects were working on task, and off when they were not. For all subjects on all days, sessions were ended only when children stated that they were ready to stop. If a subject simply stopped working, or stood looking at the experimenter, the experimenter attempted to appear busy. Eye contact was avoided, and at no time were subjects questioned as to whether they were ready to go back to the room. When the children indicated their intention to stop, they were asked to keep what they had done "a secret" and escorted back to their room.

Treatment session. On the treatment day, directions to the three groups diverged. Subjects in the control condition were presented with a different stack of arithmetic pages and given essentially the same instructions as in Baseline 1. For children in the reward-same condition, the directions for Baseline 1 were briefly repeated. The following was then added:

> Last time you did this many[2] pages and worked up to here.
>
> (The experimenter quickly flipped through the pages and drew a line at the point at which the child had stopped on the first set of papers. The word "prize" was written above the line.)
>
> If you work until you get to this line and do the same number of problems you did last time, you will earn a prize.

At this point, the child was shown a small box filled with balloons, rings, pencils, compasses, and magnets, none of which cost more than 5 cents. The box was then closed, placed on the experimenter's desk, and the directions continued:

> You do not have to stop where it says "prize." You can work as long as you want. When you are ready to stop, tell me, and you can choose the prize you want. When you choose your prize, you won't be able to do any more problems, so be sure you work as long as you want to. Remember, you have to work at least to where it says "prize" if you want to earn a prize.

When subjects indicated they were ready to stop, the experimenter checked to see whether they had worked at least as far as the line which had been drawn. Subjects were then allowed to choose a prize and asked "how they felt" about it. They were told that the next time they came to work, the experimenter would not have any prizes and were then taken

back to their room.

The procedure for children in the reward-more condition was identical to that used in the reward-same condition except for the following change in instructions:

> Last time you did this many pages and worked up to here (a dotted line was drawn). This time if you do these pages plus this many more, and go up to here (a dark line was drawn and the word "prize" written), you will earn a prize.

The instructions continued as for the reward-same subjects. Based on the results of an earlier pilot study, the work increment required was either three additional pages or 20% of the subject's output in Baseline 1, whichever was more.

Baseline 2. For all subjects, the instructions of Baseline 1 were briefly repeated and a third stack of arithmetic pages presented. In addition, for subjects in the reward-same and reward-more conditions, the following was added: "Remember, I told you last time that I wouldn't have any prizes with me today."

Assessment of the Experimental Manipulations

All subjects in the reward-same and reward-more groups did the number of problems necessary to earn the prize. When asked "how they felt" about the prize, virtually all answered with "fine," or something similar, and a smile. Only two subjects responded differently, one replying "O.K.," and one saying nothing at all. At no time did any of the children indicate any reluctance or embarrassment at choosing a prize, nor did any of the subjects express disappointment when reminded, at the start of Baseline 2, that the experimenter did not have any prizes. Each of the five prizes

was chosen by some subjects, with the balloon the favorite choice of the boys, and the ring most preferred by the girls.

Results

Between-Group Analyses

The Baseline 1 measures for number of problems completed, seconds on task, and seconds per problem were examined using analyses of variance. However, significant correlations were found between performance at Baseline 1 and performance during the treatment and Baseline 2 sessions for each of these variables. Therefore, in order to increase the sensitivity of treatment session and Baseline 2 analyses by reducing error variance due to individual differences, analyses of covariance, with Baseline 1 performance as the covariate, were used. (For individual comparisons, a t test recommended by Cohen & Cohen, 1975, was then employed.) If an initial test indicated that the assumption of homogeneity of regression was not met, however, such an analysis was inappropriate, and analysis of variance was used instead.

Number of problems completed. The number of problems completed is the variable of greatest interest since rewards during the treatment session were contingent upon subjects equalling (reward-same group) or excelling (reward-more group) their Baseline 1 output (see Figure 1).

Insert Figure 1 about here

Analysis of the Baseline 1 data reveals no significant main effects or interactions, indicating the groups did not differ in initial motivation. During the treatment session, the Group factor, as expected, is significant,

($F(2, 48) = 5.64$, $p < .01$. Significant main effects are also found for the SES and Sex factors with low-SES children and girls completing more problems than middle-SES subjects and boys, $F(1, 48) = 5.46$, $p < .025$, and $F(1, 48) = 8.29$, $p < .01$, respectively. There are no significant interactions. Individual comparisons[3] confirm that the reward-more group completed more problems than either the reward-same, $F(1, 57) = 4.67$, $p < .05$, or control subjects, $F(1, 57) = 9.44$, $p < .005$, but the difference between the latter two conditions is not significant. Analysis of the Baseline 2 data reveals the Group factor is again significant, $F(2, 47) = 3.57$, $p < .05$. Girls continue to complete more problems than boys, $F(1, 47) = 6.07$, $p < .025$, but there are no significant interactions. Individual comparisons reveal the reward-same group completed more problems than did control subjects, $F(1, 56) = 7.49$, $p < .01$. Reward-more subjects also completed more problems than the control group, but not significantly so; thus, the experimental hypothesis is not confirmed.

Seconds on task. There are no main effects or interactions during Baseline 1 for the seconds-on-task variable. Analysis of treatment session data reveals only the Group factor to be significant, $F(2, 47) = 4.55$, $p < .025$. Individual comparisons indicate that reward-more subjects worked longer than those in the reward-same or control conditions, $F(1, 56) = 8.53$, $p < .01$, and $F(1, 56) = 9.68$, $p < .005$, respectively. The reward-same and control groups do not differ. During Baseline 2, only the Sex factor emerges as significant, $F(1, 48) = 6.65$, $p < .025$, with girls working for longer periods than boys. There are no interactions. Once again, the performance of subjects in both reward groups exceeds that of

the controls although, for this veriable, not significantly so (reward-same $\bar{X}$ = 1,590 secs; reward-more $\bar{X}$ = 1,323 secs; control $\bar{X}$ = 1,110 secs).

Seconds per problem. Baseline 1 measures yield no main effects or interections. Anelyeis of the treetment session data reveels e significant group effect, $F(2, 47) = 9.31$, $p < .00i$, but no other msin effects or interactions. Individual comparisons indicate that subjects in the reward-more and reward-same groups worked fester then the controls, $F(1, 56) = 13.8$, $p < .001$, and $F(1, 56) = 14.12$, $p < .001$, respectively. On this variable, then, a reinforcement effect is found in both reward groups. The difference between the reward groups is not significant. During the Baseline 2 session, there are no significant main effects or interactions. The reward-same group once again turns in the best Baseline 2 performance (reward-same $\bar{X}$ = 11.1 secs; reverd-more $\bar{X}$ = 12. secs; control $\bar{X}$ = 12. secs).

Within-Group Analyses

Large individual differences are found in output, seconds on tesk, and seconds per problem when the Beseline 1 scores for all 60 subjects are split at the median. Those 30 subjects showing the greatest initiel motivation complete an average 103 problems, work for 1,120 secs, and finish a problem every 8.3 secs. On the other hand, the subjects least motivated initielly complete 34 problems on the average, work 297 secs, and require 12.6 secs to complete each problem. Those subjects highest in motivation do, of course, change somewhat from one variable to the other, although there are positive correlations among the three measures. It seems safe to say that there were wide differences in subject response to the tesk during Baseline 1 and that these may have reflected differences in intrinsic

motivation since, at that point, any experimentally-produced extrinsic motivation was equivalent for all subjects.

To assess within-group changes, subjects within each of the three groups were ranked according to Baseline 1 performance. The groups were then split at the median forming six subgroups, three low and three high in initial motivation on each variable. Table 1 presents the Baseline 1-Baseline 2 change scores for each subgroup as well as the results of paired t tests used to analyze the significance of the change.

Insert Table 1 about here

To date, overjustification theorists have not made a distinction between the two types of reward condition used here. Both groups met the primary stipulation of the overjustification hypothesis which is that subjects engage in an activity in order to obtain a reward. Therefore, the low subgroups of the reward-more and reward-same conditions were combined, as were the high subgroups. The change scores of these combined reward-low and combined reward-high subgroups were also analyzed and are reported in the table.

Number of Problems completed and seconds on task. The failure of an overjustification effect to appear is particularly evident when rewarded subjects are examined according to their initial performance levels. Those subjects who were originally below the median (combined reward-low subgroups) in output and seconds on task show a significant increase in performance from Baseline 1 to Baseline 2, while rewarded subjects who were originally most highly motivated (combined reward-high subgroups) evidence no decline. On the other hand, no change is seen in controls initially

low in motivation, while control subjects who were highly motivated in Baseline 1 show a significant decline in output and seconds on task by Baseline 2. The fact that the control subjects initially lowest in output did not also significantly decline across sessions may be due to a floor effect, as 7 of the 10 subjects in this group decreased in output despite their already low performance level. By way of comparison, only 5 of the 20 subjects in the combined reward-low subgroup showed such a decroase.

Seconds per problem. Significant changes in this variable are found only if the reward groups are considered separately. Reward-more subjects who were slowest initially, and reward-same subjects who were fastest initially, both show a significant decline in performance from Baseline 1 to Baseline 2. This is the only indication in the study of significant decreases in motivation following reward withdrawal. It should be noted, however, that subjects in the reward-same (high) subgroup worked more quickly than the control (high) subgroup during Baseline 2 ($\bar{X}$ = 8.7 secs and $\bar{X}$ = 10.3 secs, respectively). Moreover, in both the reward-same (high) and control (high) subgroups, seven subjects were working more slowly, and three subjects more quickly, by Baseline 2. Whatever the reason for the decline in performance found, when the reward-low subgroups are combined, and the reward-high subgroups are combined, the significant Baseline 1-Baseline 2 differences disappear.

Accuracy and Number of Problems Skipped

The mean level of accuracy in Baseline 1 was 97%, confirming that the task was one for which the subjects had achieved a high level of mastery. A measure of accuracy was included for several reasons. For control purposes, it was important to ascertain that there were no significant

SES or sex differences in terms of ability to do the problems, since this could conceivably affect motivation. The data confirm that subjects of both economic groups, and both sexes, were consistently and highly accurate (low-SES $\bar{X}$ = 97%; high-SES $\bar{X}$ = 98%; male $\bar{X}$ = 98%; female $\bar{X}$ = 97%).

Aside from issues of control, however, it is of particular interest to examine accuracy levels during the treatment session. The advantages of having persons produce more, spend more time on task, or work faster, would be offset if accuracy declined. The data indicate that accuracy is not detrimentally affected by either reward introduction or reward withdrawal. During the treatment session, subjects in the reward groups averaged 97.5%, while during the Baseline 2 period, they performed at a 97.9% level of accuracy.

A second check on quality of performance was made by measuring the number of problems skipped by subjects. It was thought that children in the reward groups might work more carelessly during the treatment session and perhaps skip problems in order to gain the prize more quickly. The data indicate, however, that subjects in all groups worked carefully in this regard during each of the sessions. On the average, subjects in reward groups, as well as the control group, skipped one problem during the treatment session. During Baseline 2, rewarded subjects averaged one skipped problem, compared to two problems for controls. There is again no evidence that reward withdrawal had a detrimental effect on quality of performance through some type of "let down" factor. In sum, it appears that the quantitative gains in output, seconds on task, and seconds per problem, were not earned at the expense of performance quality.

Discussion

The overall picture that emerges from the results obtained is that rewards played a significant role in maintaining or enhancing intrinsic motivation which otherwise, in the absence of reward, might have deteriorated. On each quantitative variable, the Baseline 2 performance of the reward-same subjects was the highest of the three groups, while that of the control subjects was the lowest. For the output variable, this difference was significant. Thus, the overjustification hypothesis, which predicts significantly poorer performance, compared to controls, when subjects engage in an activity to attain a reward, was not supported by the data. Such a result is all the more striking since the experimental procedures were deliberately set up so as to satisfy all the conditions thought to be particularly deleterious to intrinsic motivation. Thus, the task was already well-mastered (cf. Lepper & Greene, 1976); the rewards were salient (Ross, 1975), contingent and expected (Lepper, Greene, & Nisbett, 1973), noncommensurate, i.e., not matched to a specific subject need (Eden, 1975), and for a single trial (Reiss & Sushinsky, 1975). In addition, the basic conditions required for the overjustification effect were satisfied: The activities were of at least some interest (particularly to those subjects above the median on Baseline 1 measures of performance); subjects did not expect further extrinsic rewards in Baseline 2; the instrumentality of engaging in the task to obtain the rewards was made salient. In fact, the experimental procedure went one step further during the treatment session by emphasizing to the subjects in the reward-same group that they had done the same amount of work previously when unrewarded.

Although the hypothesis that the Baseline 2 output of the reward-more

group would significantly exceed that of the other two groups was not confirmed, it should be noted that the difference in output between the reward-more and control groups was in the expected direction, and this was also true for the seconds-on-task variable. (There was no difference in performance on the seconds-per-problem variable.) This, coupled with the consistently superior Baseline 2 performance of the reward-same group, provides at least some support for reinforcement theory. It is possible that satiation or fatigue depressed the reward-more group's performance during Baseline 2 inasmuch as they significantly outperformed the other two groups during the treatment session.

The fact that both reward groups performed significantly better than the control group during the treatment session bears on an issue concerning the production of reinforcement effects in overjustification experiments. Lepper and Greene (1976, p. 31) have maintained that it is "of considerable theoretical interest" that overjustification effects "can be produced without a prior reinforcement effect." However, Reiss and Sushinsky (1975, 1976) have argued that unless a procedure is first demonstrated to produce a reinforcement effect, a decrease in motivation might be due to aversive procedures or competing responses such as distraction or frustration, as well as overjustification. This experiment and a study by Moracco and Fashoh (1978) are apparently the first to report a reinforcement effect while using an overjustification paradigm which included a control group. That neither found an overjustification effect is in accord with tho prediction of Reiss and Sushinsky.

Some comments regarding control group performance, which consistently dropped from Baseline 1 to Baseline 2, are in order. It is possible that rather than being a neutral condition of no reward, the control condition

was actually aversive. There were no indications that this was the case, however, as all subjects came quite willingly, and any comments made regarding the task were uniformly positive.

Another possibility is that lack of feedback had detrimental effects on the control group's motivation. None of the groups received explicit feedback regarding their performance, but the reward groups could have interpreted the contingency instructions as an indirect form of feedback regarding their Baseline 1 performance, as well as goal setting for the treatment session. Although the goal-setting instructions led to higher reward-group performance during the treatment session, overjustification theorists would expect such procedures to produce motivational decrements following reward withdrawal. This is because the surveillance and evaluation implied in feedback and goal setting are viewed as forms of extrinsic control. (See Lepper & Greene, 1975, for an examination of the negative consequences of surveillance; see Maehr, 1976, for a discussion of evaluation effects.)

Socioeconomic Status and Sex Differences

Low-SES subjects consistently completed more problems, spent more time on task, and worked more quickly than did the middle-SES subjects, although the output variable was the only one for which the difference was significant. Ethnicity and SES were confounded, but this fact does not provide any ready explanation for the differences found. Children at the low-SES school were receiving a type of reading instruction which requires a high degree of concentration for 20-minute time periods, and this may have affected their performance. Although there was a main effect for SES, there were no significant Group x SES interactions, indicating

that, regardless of social class, children responded similarly to the experimental conditions.

Girls consistently completed more problems and worked for longer periods than the boys. This main effect for sex may have reflected the generally more mature behavior of the six-year-old girl as compared to her male peer (cf. Watson & Lindgren, 1973), or it may be that girls of this age are more responsive to social situations, particularly those involving an adult female. That there were no Group x Sex interactions is in agreement with the results of several previous studies (cf. Greene & Lepper, 1974; Lepper, Greene, & Nisbett, 1973; Ross, Karniol, & Rothstein, 1976).

Within-Group Changes

The most theoretically interesting results involved within-group changes. These can be seen most clearly by examining Figure 2 which presents the results obtained when the two levels of extrinsic reward--no reward (control) and (combined) reward groups--are considered jointly with the two levels of intrinsic reward--subjects initially low or high on Baseline 1 measures. (The results were identical for the measures of output and seconds on task; there were no significant differences on the seconds-per-problem variable.)

Insert Figure 2 about here

Again, it can be seen that subjects in the High Intrinsic Interest-Reward cell did not show the decrease in motivation predicted by the overjustification hypothesis. The Low Intrinsic Interest-No Reward

cell represents a condition which has been extensively explored in the research on cognitive dissonance. Although it has been suggested that such a condition may lead to an increment in intrinsic motivation (cf. Deci, 1975, chap. 6), the present study provides no evidence that motivation is enhanced in such a condition.

The remaining two conditions are of interest because they represent situations often found in applied settings. Thus, persons who have little interest in certain activities are sometimes offered incentives if they will improve, or at least maintain, their level of performance. This is usually the case when token economies, represented here by the Low Intrinsic Interest-Reward cell, are instituted. Some overjustification theorists maintain the results of their work suggest that token programs may, in the long run, have harmful effects on intrinsic motivation (cf. Lepper & Greene, 1976). For the present task, rewarding subjects who had little initial interest in the activity led to significantly elevated levels of performance which were maintained even after rewards were withdrawn.

On the other hand, it is more common than one might suppose for an individual to perform at a relatively high level and receive little reward or even feedback (High Intrinsic Interest-No Reward condition). It is not at all unusual, for instance, for students to work three days, as did the control subjects of this experiment, without receiving information regarding their performance. It is even less unusual for students to work for such a period without receiving some sort of positive reinforcement, whether in the form of praise or something more tangible. The precipitous decline in the performance of control subjects who were above the median on Baseline 1 measures suggests this may have a particularly

deleterious effect on subjects who are highly motivated initially. Such a result is not predicted by overjustification theorists who seem to regard intrinsic motivation as fairly stable unless detrimentally affected by the introduction of reward or some other form of external constraint. One plausible explanation for such a decline is suggested by a study (Cherrington, Reitz, and Scott, Jr., 1971) which found that high performers may feel cheated when they do not receive contingent reward, and that such feelings may lead to dissatisfaction and eventual dropping-out behavior.

The concept of payment norms may also explain the failure to find an overjustification effect in the High Intrinsic Interest-Reward group. When Staw, Calder, and Hess (Note 1) manipulated norms for payment, they found that reward decreased intrinsic interest only when there existed a situational norm for no payment. With reference to the present study, it seems reasonable that prizes would appear more appropriate to children who have been working on arithmetic problems than they would to children in earlier studies who have been rewarded for drawing with magic markers or playing with a drum. Research to investigate the realistic-unrealistic task dimension is needed. Since the present data confirm that a realistic task can be highly motivating for many subjects, such tasks need not be ruled out for use in experimentation in this area. Furthermore, if realistic tasks are used, there would be a firmer basis for generalizing results to natural settings.

A note: Researchers have, with few exceptions, simply assumed that certain tasks were highly interesting to all subjects. However, the data presented indicate that there may be large individual differences in the extent to which subjects find tasks interesting. It therefore seems prudent for future investigators to furnish some validation of task

interest levels, given the importance assigned to such levels by the overjustification hypothesis.

Too Much Reward--Or Too Little?

Despite the comments of those who have spoken against the "plethora of extrinsic rewards" in schools today (Deci, 1975, p. 212), a case can be made that many students go unrewarded much of the time. Researchers investigating the ratio of positive to negative comments in classrooms rather consistently report a preponderance of unfavorable remarks. Although many have expressed concern about the possibly negative effects of overreward on motivation, virtually no attention has been paid to the possibly negative consequences of underreward. Given the rapid deterioration of performance found here when subjects high in motivation went unrewarded, the phenomenon of undercompensation and its effects seems worthy of further investigation. Also valuable would be additional research to determine the ways in which the negative effects identified by overjustification studies can be avoided, neutralized, or even reversed. Procedures which emphasize to individuals that their behavior is under their own control, multiple trials of reward, and phasing-out-of-reward strategies are all just beginning to be explored. Overjustification research is a valuable line in inquiry in that it points out problems which may occur when rewards are employed. The results of this study emphasize the importance of seeking solutions.

Reference Note

1. Staw, B. M., Calder, B. J., & Hess, R. K. Intrinsic motivation and norms about payment. Unpublished manuscript, Northwestern University, 1975.

References

Anderson, R., Manoogian, S. T., & Reznick, J. S. The undermining and enhancing of intrinsic motivation in preschool children. Journal of Personality and Social Psychology, 1976, 34, 915-922.

Arnold, H. J. Effects of performance feedback and extrinsic reward upon high intrinsic motivation. Organizational Behavior and Human Performance, 1976, 17, 275-288.

Bem, D. J. An experimental analysis of self-persuasion. Journal of Experimental Social Psychology, 1965, 1, 199-218.

Bem, D. J. Self-perception: An alternative interpretation of cognitive dissonance phenomena. Psychological Review, 1967, 74, 183-200.

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 6). New York: Academic Press, 1972.

Calder, B. J., & Staw, B. M. Self-perception of intrinsic and extrinsic motivation. Journal of Personality and Social Psychology, 1975, 31, 599-605.

Cherrington, D. J., Reitz, H. J., & Scott, Jr., W. E. Effects of contingent and noncontingent reward on the relationship between satisfaction and task performance. Journal of Applied Psychology, 1971, 55, 531-536.

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. New York: Halsted Press, 1975.

Davis, R. T., Settlage, P. H., & Harlow, H. F. Performance of normal and brain-operated monkeys on mechanical puzzles with and without food incentive. The Journal of Genetic Psychology, 1950, 77, 305-311.

de Charms, R. Personal causation. New York: Academic Press, 1968.

Deci, E. L. Intrinsic motivation, extrinsic reinforcement, and inequity. Journal of Personality and Social Psychology, 1972, 22, 113-120.

Deci, E. L. Intrinsic motivation. New York: Plenum Press, 1975.

Deci, E. L., Cascio, W. F., & Krusell, J. Cognitive evaluation theory and some comments on the Calder and Staw critique. Journal of Personality and Social Psychology, 1975, 31, 81-85.

Eden, D. Intrinsic and extrinsic rewards and motives: Replication and extension with kibbutz workers. Journal of Applied Social Psychology, 1975, 5, 348-361.

Greene, D., & Lepper, M. R. Effects of extrinsic rewards on children's subsequent intrinsic interest. Child Development, 1974, 45, 1141-1145.

Greene, D., Sternberg, B., & Lepper, M. R. Overjustification in a token economy. Journal of Personality and Social Psychology, 1976, 34, 1219-1234.

Hodges, W. L. The role of rewards and reinforcements in early education programs: I External reinforcement in early education. Journal of School Psychology, 1972, 10, 233-241.

Kruglanski, A. W., Alon, S., & Lewis, T. Retrospective misattribution and task enjoyment. Journal of Experimental Social Psychology, 1972, 8, 493-501.

Kruglanski, A. W., Riter, A., Amitai, A., Margolin, B., Shabtai, L., & Zaksh, D. Can money enhance intrinsic motivation?: A test of the content-consequence hypothesis. Journal of Personality and Social Psychology, 1975, 31, 744-750.

Lepper, M. R., & Greene, D. Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. Journal of Personality and Social Psychology, 1975, 31, 479-486.

Lepper, M. R., & Greene, D. On understanding "overjustification": A reply to Reiss and Sushinsky. Journal of Personality and Social Psychology, 1976, 33, 25-35.

Lepper, M. R., Greene, D., & Nisbett, R. E. Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. Journal of Personality and Social Psychology, 1973, 28, 129-137.

Maehr, M. L. Continuing motivation: An analysis of a seldom considered educational outcome. Review of Educational Research, 1976, 46, 443-462.

Moracco, J. C., & Fasheh, V. Effects of contingency management on academic achievement and conduct of mentally retarded Arab students. American Journal of Mental Deficiency, 1978, 82, 487-493.

Notz, W. W. Work motivation and the negative effects of extrinsic rewards. American Psychologist, 1975, 30, 884-891.

Pierce, R. E. The effect of monetary rewards on improved academic performance (Doctoral dissertation, University of Southern Mississippi, 1970). Dissertation Abstracts International, 1971, 31, 3885A. (University Microfilms No. 71-5398, 71)

Reiss, S., & Sushinsky, L. W. Overjustification, competing responses, and the acquisition of intrinsic interest. Journal of Personality and Social Psychology, 1975, 31, 1116-1125.

Reiss, S., & Sushinsky, L. W. The competing response hypothesis of decreased play effects: A reply to Lepper and Greene. Journal of Personality and Social Psychology, 1976, 33, 233-244.

Ross, M. Salience of reward and intrinsic motivation. Journal of Personality and Social Psychology, 1975, 32, 245-254.

Ross, M., Karniol, R., & Rothstein, M. Reward contingency and intrinsic motivation in children: A test of the delay of gratification hypothesis. Journal of Personality and Social Psychology, 1976, 33, 442-447.

Sorensen, R. L., & Maehr, M. L. Toward the experimental analysis of "continuing motivation." The Journal of Educational Research, 1976, 69, 319-322.

Watson, R. I., & Lindgren, H. C. Psychology of the Child (3rd ed.). New York: Wiley, 1973.

Footnotes

The author wishes to thank Carl Spring for his suggestions and assistance at all stages of this research. Stanley Coopersmith and Tom Klein provided valuable comments on an earlier draft.

[1]Only studies which included a nonrewarded control group and yielded significant results ($p < .05$ or less) will be cited.

[2]Subjects were not told the number of pages completed during Baseline 1 to help prevent their stopping on a particular page during the following sessions simply because they had stopped at that point before.

[3]All p values are based on two-tailed tests of significance.

Table 1

Within-Group Analyses for Subjects Above and Below the Median on Baseline 1 Performance for the Variables of Output, Seconds on Task, and Seconds Per Problem

	Output		Seconds on Task		Seconds Per Problem	
Group	Change Score[a]	t	Change Score	t	Change Score	t
Reward-More[b]						
Low	+12.8	n.s.	+212.0	2.66*	-2.44	3.44***
High	-20.7	n.s.	-162.0	n.s.	-1.22	n.s.
Reward-Same						
Low	+13.8	n.s.	+108.0	n.s.	+ .57	n.s.
High	+18.0	n.s.	+174.0	n.s.	- .81	2.77*
Control						
Low	+ 1.3	n.s.	+ 54.8	n.s.	-1.86	n.s.
High	-65.7	4.38***	-617.5	4.17***	-1.73	n.s.
Combined Reward-Low Subgroups[c]	+26.6	2.23*	+320.0	3.33***	-1.87	n.s.
Combined Reward-High Subgroups	- 2.7	n.s.	+ 12.0	n.s.	-2.03	n.s.

[a]Change score = Baseline 2 - Baseline 1 performance.

[b]Each group is divided into those subjects lowest and highest on Baseline 1 measures. n = 10 in each subgroup.

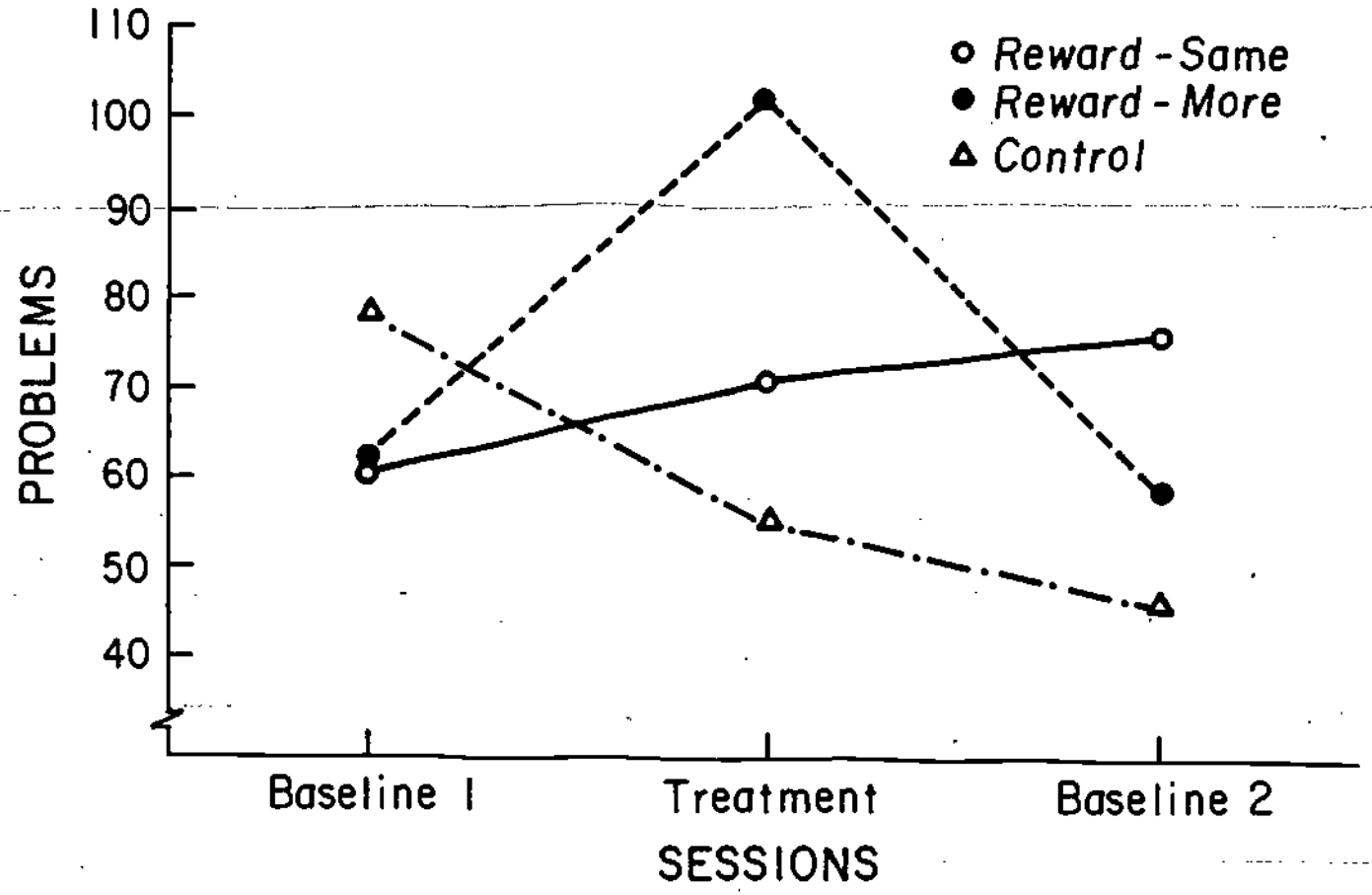
[c]n = 20 in combined reward-subgroups.

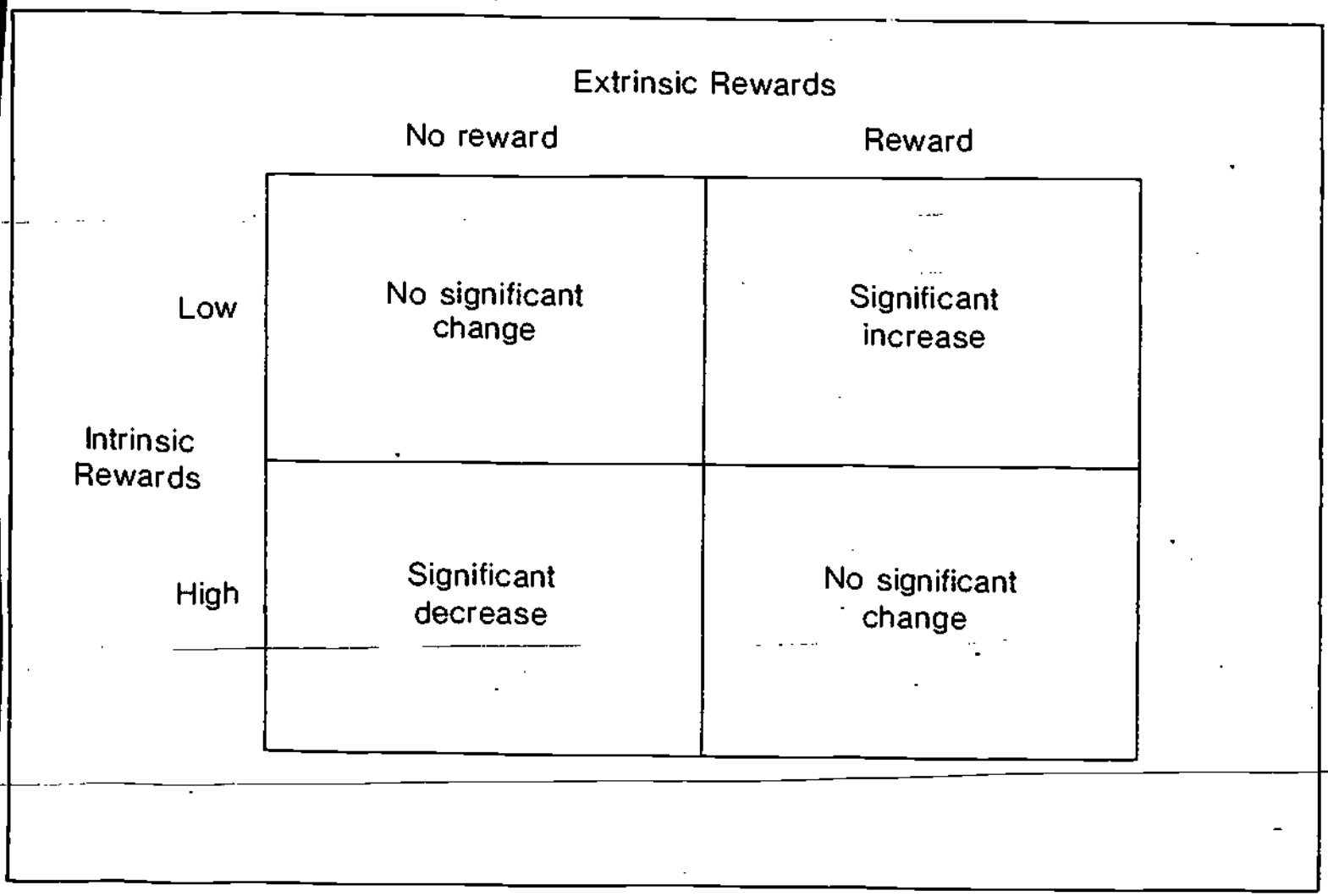
*$p < .05$ **$p < .02$ ***$p < .01$.

Figure Captions

Figure 1. Mean number of problems completed by each group during each session.

Figure 2. Comparison of Baseline 1 and Baseline 2 measures of output and seconds on task.


110
100
90
80
70
60
50
40
PROBLEMS
Reward - Same
Reward - More
Control
Baseline 1
Treatment
Baseline 2
SESSIONS


Extrinsic Rewards
No reward
Reward
Intrinsic Rewards
Low
High
No significant change
Significant increase
Significant decrease
No significant change