

Effects of No Feedback, Task-Related Comments, and Grades on Intrinsic Motivation and Performance

Ruth Butler and Mordecai Nisan
School of Education
Hebrew University of Jerusalem, Jerusalem, Israel

This study was designed to test the hypothesis that intrinsic motivation would be maintained after receipt of nonthreatening, task-related evaluation and undermined after repeated non-receipt of feedback or receipt of controlling normative grades. Nine classes comprising 261 sixth-grade pupils were randomly assigned to one of these three feedback conditions and were given two interesting tasks, one quantitative and one qualitative, on three sessions over 2 days. The manipulation was applied after Sessions 1 and 2, and no feedback was expected or received after Session 3. Experimental measures consisted of Session 3 performance scores and of the results of a questionnaire, given after Session 3, which tapped interest and patterns of attribution of success and effort. The results confirmed the hypothesis and revealed significant group differences in intrinsic motivation as reflected in both performance and attitudes.

The need for mastery has received considerable attention in recent years, with different researchers presenting what can be viewed as different aspects of this need (DeCharms, 1968; Deci, 1975; Harter, 1981). One aspect that merits further study is the search for information about one's competence and success in a task (Festinger, 1954; Suls & Miller, 1977). Such information seems vital to a sense of mastery and self-determination because without it one cannot assess one's mastery in any given task. Thus one would expect the availability (and/or expectation) of feedback to be an important factor in task motivation in general and in determining interest, or intrinsic motivation, in particular. Specifically, one would expect intrinsic motivation to be greater for tasks perceived as supplying information about competence and to be undermined when no such information is expected.

Similar arguments have been raised, or implied, in some of the literature concerned with the effects of extrinsic incentives on intrinsic motivation (e.g., Deci, 1975). Deci's cognitive evaluation theory argues that rewards have two aspects — a controlling one that encourages attribution of behavior to extrinsic factors and thus undermines the individual's sense of self-determination, and an informational one compatible with needs for competence and control. When the former aspect is salient, intrinsic motivation will be undermined; when positive information about efficacy is prominent, the sense of mastery will be enhanced, and subsequent interest maintained, or even increased. However, as formulated by Deci, the hypothesized relationship seems in fact to be that between self-enhancement and task motivation. This is also reflected in several studies that hypothesized, and found, that receipt of positive informa-

tion about task performance tended to sustain interest at a level similar to that of controls who received no information, whereas receipt of negative feedback undermined subsequent interest (Boggiano & Ruble, 1979; Deci, Cascio, & Krusell, 1973). Similarly, Deci (1971) found that verbal praise increased later interest relative to that of controls.

Although such studies demonstrate the importance of the need for self-enhancement, no studies have as yet directly tested the role of the availability of diagnostic information relevant to self-assessment of competence on continuing task motivation. Even Pittman, Davey, Alafat, Vetherill, and Kramer (1980), who reported finding that informative feedback enhanced intrinsic motivation, may in fact have been studying the effects of generalized verbal praise ("You are doing very well") rather than of specific task-related feedback.

The present research was thus designed to study the effects of different feedback conditions on intrinsic motivation. We expected that receipt of individualized, specific, non-normative information about task performance, including both positive and negative comments, would maintain or even enhance subsequent motivation. The role of the availability of such information was studied in comparison with conditions of nonreceipt of any information and of receipt of normative evaluation. Most studies of extrinsic incentives and intrinsic motivation, including those mentioned earlier, used as controls subjects who received no rewards or feedback, apparently on the assumption that under these conditions original levels of intrinsic motivation would be maintained. Although this may be true within the one-trial experimental paradigm favored by most researchers, recognition of the centrality of self-assessment needs suggests that interest in a task that repeatedly fails to satisfy these needs will eventually wane. Finally, the normative grades prevalent in many schools seem to provide a striking example of feedback in which control rather than information is salient. As such, we would expect them to undermine intrinsic motivation. This expectation receives some

We would like to thank Itamar Gati and Sorel Cahan for their helpful comments on an earlier draft of this article.

Correspondence concerning this article should be addressed to Ruth Butler, School of Education, Hebrew University of Jerusalem, Jerusalem, Israel 91905.

support from Harter's (1978) finding that expectation of letter grades affected children's task motivation in ways similar to the various extrinsic rewards used in other studies.

The motivational effects predicted above were tapped by measures of both performance and attitudes. Several studies reviewed, for example, by Condry (1977) and McGraw (1978) suggest that although extrinsic motivational conditions may result in superior performance on quantitative, speed, or algorithmic tasks, performance on qualitative, power, or heuristic tasks involving problem solving or divergent thinking is maintained under intrinsic conditions but undermined under extrinsic ones. Thus the effects of the manipulation on performance were studied by using both a quantitative "speed" task and a qualitative "power" one. Expressed interest and willingness to perform further tasks and patterns of attribution of effort and success were the attitudinal measures used. The effects of each motivational condition on these measures was studied by using a three-trial design. The specific hypotheses tested can be formulated as follows: (a) Subjects who receive individualized task-related feedback contingent on the performance of interesting tasks over two trials will subsequently express more interest in the tasks than will subjects who receive normative numerical grades or no feedback. (b) Performance on the "qualitative" task will be higher after receipt of task-related evaluation than after receipt of numerical grades or no feedback. (c) Performance on the "quantitative" task will be higher after receipt of task-related evaluation or numerical grades on two previous trials than after repeated nonreceipt of feedback.

Method

The hypotheses were tested by using a mixed design. Subjects received a similar pair of tasks at each of three sessions. After each of the first two sessions, Group 1 received task-related written comments on their performance, Group 2 received numerical grades, and Group 3 received no evaluation. An attitudinal questionnaire was given after performance of the tasks at Session 3.

Subjects

The sample comprised 261 sixth-grade children (145 girls, 116 boys) with a mean age of 12.3 years. The children attended nine classes in three city elementary schools serving predominantly middle-class populations. Three classes were randomly assigned to each experimental group, yielding 88 subjects in Group 1, 90 in Group 2, and 83 in Group 3.

Instruments

The instruments consisted of three work booklets containing the experimental tasks, which were administered at Sessions 1, 2, and 3, respectively. In addition, a questionnaire designed to measure overt motivational attitudes was administered at the end of Session 3. Each booklet contained two tasks, A and B. In Task A of Booklets 1 and 3 (for Sessions 1 and 3, respectively) children were requested to construct as many words as they could from the letters of a longer word. Task B consisted of two examples from

the divergent thinking "uses" test (Torrance & Templeton, 1963). The tasks for Session 2 were slightly different in order to reduce boredom and practice effects. In one task children were asked to construct a word tree using the first and last letters of each preceding word, and the other task consisted of the "circles" test (Torrance & Templeton, 1963). A pilot study established that sixth-grade children found the experimental tasks interesting and that the tasks given at Sessions 1 and 3 yielded equivalent levels of performance.

The first two questions on the motivation questionnaire (see Table 2) tested expressed interest in the experimental tasks; further questions tapped attribution of effort to various factors and perceptions of the factors determining success in the tasks. Finally, pupils were asked to express a preference for one of the three modes of evaluation employed in the study. Most questions were answered on a continuous 7-point scale (1 = *low agreement* and 7 = *high agreement* with the item in question). Other questions were presented in multiple choice form.

Procedure

The experiment consisted of three sessions. Session 1 was conducted in 1 day, and Sessions 2 and 3 were conducted 2 days later, with an interval of 2 hr between them.

The experiment was conducted in each class during regular school hours by one of two female graduate students in psychology. In Session 1 instructions were identical for each group and were printed in Booklet 1. It was explained that the experimenters had constructed some tasks and needed to see how different children answered them; they hoped that the children would enjoy doing them. Instructions for Task A were then read out loud. These included rules regarding acceptable words and criteria for successful performance (according to the number and length of words). Children were asked to begin, and after 5 min they were asked to stop and turn to Task B. Instructions for Task B, adapted from Torrance and Templeton (1963), also included criteria for success (according to the number, variety, and originality of responses). After 5 min the booklets were collected.

In Session 2, 2 days later, Booklet 1 was returned. Children in Group 1 were told that each had been given an appropriate evaluation of his or her performance. Children in Group 2 were told that each had been given a grade, and those in Group 3 were told that the booklet was being returned to them. Subjects were instructed to look through the booklet to see how they had done, after which they would be given new tasks. Children in Group 1 were told that they would receive comments on the new tasks too, children in Group 2 that they would receive grades, and children in Group 3 that this booklet would also be returned. Booklet 2 was then given out, and the procedure for Tasks A and B, including repetitions of the criteria for successful performance, was followed as described for Booklet 1.

In Session 3, 2 hr later, Booklet 2 was returned with the evaluation appropriate to each group, and Booklet 3 was distributed. Subjects were given a few minutes to look through Booklet 2 and were then told that the experimenters had some tasks that had not yet been tried out. The children were asked to try out the tasks and were told that this time the experimenters would not be able to return and/or to evaluate their work. The procedure for Tasks A and B was then identical to that for previous trials. The attitude questionnaire was then distributed and the 7-point scoring system demonstrated.

Evaluation

Group 1 (task-related comments). Evaluation consisted of one sentence, which related specifically to the performance of the

Table 1
Means and Standard Deviations for Performance Scores at Session 1 and Session 3 for Three Feedback Conditions

Score	Comments group		Grades group		No feedback group	
	Session 1	Session 3	Session 1	Session 3	Session 1	Session 3
Task A						
No short words						
<i>M</i>	10.14	17.03	10.41	17.81	9.78	13.86
<i>SD</i>	4.18	6.48	6.10	7.23	6.08	8.23
No long words						
<i>M</i>	5.10	6.76	5.46	5.89	5.00	2.64
<i>SD</i>	2.39	3.11	4.89	5.13	3.00	2.30
Final score						
<i>M</i>	38.68	55.49	39.84	52.59	39.95	29.46
<i>SD</i>	14.86	19.26	24.43	25.32	17.79	14.00
Task B						
Fluency						
<i>M</i>	11.93	13.91	11.42	10.66	11.08	8.73
<i>SD</i>	4.33	5.14	4.65	4.81	4.20	4.23
Flexibility						
<i>M</i>	7.13	8.70	8.08	4.18	7.90	4.29
<i>SD</i>	2.77	3.68	2.74	3.39	3.72	3.00
Elaboration						
<i>M</i>	2.08	3.53	2.27	0.96	1.60	0.75
<i>SD</i>	2.12	2.05	2.16	1.48	1.75	1.27
Originality						
<i>M</i>	1.73	2.74	1.61	0.66	1.54	0.48
<i>SD</i>	1.21	1.38	1.13	0.86	1.19	0.72
Final score						
<i>M</i>	24.93	32.59	25.04	17.08	24.19	15.06
<i>SD</i>	9.66	11.65	8.15	8.61	9.63	8.04

individual child. It consisted of one phrase relating to some aspect of the task that the child had performed well and one phrase relating to an aspect performed less well. It did not include any information beyond that specified in the criteria for success given in the general instructions. Typical sentences for Task A were as follows (translated from Hebrew): "The words you wrote were correct, but you did not write many words." "You wrote many short words, but not many long ones." For Task B the following sentences were typical: "You thought of many ideas, but not many unusual ones." "Your ideas were very unusual, but you did not think of many ideas."

Group 2 (numerical grades). Scores were computed according to the criteria for success outlined in the instructions (see next section). Grades were then computed and awarded so as to follow a normal distribution ranging from 30 to 100 within each class.

Scoring

Task A. Scoring was according to the criteria for successful performance given in the instructions. Two-letter words were given 2 points; three-letter words, 3 points; four-letter words, 5 points; and words of five or more letters, 6 points. Note was also made of the number of short (two- or three-letter) and long (four- or more letter) words produced.

Task B. Scoring was according to the categories defined by Torrance and Templeton (1963). One point was given for each response (fluency), category (flexibility), and elaborated response, and 2 points were given for each original response, defined as a response appearing in no more than 10% of protocols. A final

score was computed from the sum of scores in each category. Scores represent average scores for each category over the two examples of the "uses" test given in each session.

Results

Performance Measures

Table 1 presents group means and standard deviations for each measure of performance at Session 1 and Session 3. First, performance scores from Session 1 were analyzed using a one-way analysis of variance (ANOVA) to check initial equivalence of the experimental groups. The analysis revealed that only for one variable, elaboration, was the group effect considerable, $F(2, 258) = 2.56, p < .08$ whereas for all other variables the obtained F was very low. The hypotheses were then tested using analyses of covariance (ANCOVAs) performed for the various measures of performance from Session 3, using the corresponding scores from Session 1 as covariants. Because preliminary analyses indicated no significant main or interaction effects for sex on any measures, data for the sexes were combined.

Task A. The results indicated a significant effect of the manipulation for the final score, $F(2, 257) = 77.00, p < .001$. The group factor was then partitioned into orthogonal planned contrasts to test the specific hypothesis that performance on Task A would be higher after receipt of comments or grades than after nonreceipt of any feedback (Groups

Table 2
Means and Standard Deviations for Experimental Groups on Attitudinal Items

Item	Comments group		Grades group		No feedback group	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
How interesting were the tasks?	6.26	.965	5.27	1.82	4.99	1.33
How many more would you like to receive?	5.65	1.01	3.44	1.31	3.71	1.27
Attribution of effort to:						
Interest	6.77	1.41	5.62	1.89	5.73	1.73
Importance of success	5.47	1.82	5.78	1.42	5.35	1.35
Influence on grades	2.84	2.12	3.06	1.93	3.35	1.87
Avoidance of poor achievement	5.02	1.78	5.43	1.23	4.75	2.13
Attribution of success to:						
Skill	5.63	1.65	4.46	2.09	5.27	1.81
Pupil's mood	5.39	2.10	4.56	2.38	5.19	2.07
Effort	6.23	1.57	6.01	1.36	5.53	1.95
Reader's mood	2.30	2.01	2.99	2.50	4.45	2.68
Interest	6.45	1.36	5.73	1.78	5.81	1.58
Neatness	2.06	1.76	3.53	2.20	4.98	2.19

1 and 2 vs. Group 3), whereas the scores of Groups 1 and 2 would be similar. As expected, the former contrast was highly significant, $t(1, 257) = 12.30, p < .001$, and no significant difference was found between Groups 1 and 2. As shown in Table 1, final scores at Session 3 were similar in the comments and grades groups, which in turn were higher than in the no-feedback group. The same significant trends were also obtained for the components of Task A: $F(2, 257) = 9.24, p < .001$, for short words, and $F(2, 257) = 43.25, p < .001$, for long words. The Session 3 scores of the two feedback groups were again higher than those of the no-feedback group: $t(1, 257) = 4.25, p < .001$, for short words, and $t(1, 257) = 8.97, p < .001$, for long words. For the number of long words produced, the comparison between the grades and comments groups was also significant, $t(2, 257) = 2.50, p < .013$, with the comments group scoring higher than the grades group. Thus the hypothesis that performance on the quantitative task would be higher after receipt of some feedback than after nonreceipt of feedback was supported by the results.

Task B. It was hypothesized that performance on this "qualitative" task would be higher after receipt of comments than after receipt of grades or no feedback. The results of the ANCOVA revealed a significant effect of the manipulation for final scores on Task B, $F(2, 257) = 123.28, p < .001$. Specific hypotheses were tested using orthogonal planned contrasts comparing the comments group with the grades and no-feedback groups, and the grades group with the no-feedback group. As predicted, the comments group scored significantly higher than the others, $t(1, 257) = 15.68, p < .001$, and there was no significant difference between the grades and no-feedback groups (see Table 1). The same significant main and contrast effects were also obtained for all the subscores of Task B: for fluency, $F(2, 257) = 31.00$ and $t(1, 257) = 7.40$; for flexibility, $F(2, 257) = 62.88$ and $t(1, 257) = 11.20$; for elaboration, $F(2, 257) = 91.88$ and $t(1, 257) = 13.55$; and for originality, $F(2, 257) = 160.01$ and $t(1, 257) = 17.88$. All results

were significant at $p < .001$. The comparison between the grades and no-feedback groups was significant only for the fluency component of Task B, $t(2, 257) = 2.88, p < .004$, with the grades group scoring higher (see Table 1).

Verbal Measures

The means and standard deviations for the responses of each group to items on the motivation questionnaire are presented in Table 2. In view of the practical difficulties involved in repeated answering of similar questions, no formal attempt to establish the reliability of these items was made. However, an estimate for the lower bound of reliability can be found in the high correlation ($r = .81$) between the two similar, but not identical, questions tapping interest.

The motivational data were analyzed using a one-way ANOVA. When the obtained F was significant, Scheffé's test with a significance level of $p < .05$ was run to test for differences in the means of the three experimental groups.

Measures of interest. The hypothesis that the manipulation would affect expressed interest in the tasks was supported by the results, $F(2, 258) = 25.48, p < .001$. Scheffé's test revealed that subjects who received comments expressed significantly more interest than did those who received grades or no feedback. A similar effect was found for the willingness of subjects to volunteer for further tasks, $F(2, 258) = 46.59, p < .001$.

Attribution of effort. Pupils were asked to attribute effort to each of four factors in turn. A significant group effect was received only for attribution to interest, $F(2, 258) = 20.87, p < .001$, with the comments group scoring significantly higher than the other groups. Intergroup differences were more marked in response to a question asking which of the four factors was most influential in determining the effort invested in the tasks. The percentages of subjects choosing each factor in each group are presented in Table

3. The comments group overwhelmingly chose interest. The grades group also frequently chose importance and the desire to avoid poor achievement, whereas the no-feedback group tended to attribute effort equally to interest and importance. A chi-square analysis revealed a highly significant association between experimental group and the choice of a determinant of effort, $\chi^2(6, N = 261) = 35.72, p < .001$.

Attribution of success. The ANOVAs revealed a significant effect of the manipulation for all factors. The effect on attribution to skill, $F(2, 258) = 9.49, p < .001$, derived from the tendency revealed by Scheffé's test for the comments and no-feedback groups to attribute success to skill significantly more than did the grades group. A similar pattern accounted for the significant effect of group on attribution of success to pupil's mood, $F(2, 258) = 3.38, p < .04$. The comments group also tended to attribute success more to effort, $F(2, 258) = 4.04, p < .02$, and to interest, $F(2, 258) = 5.50, p < .005$, than did pupils in the other groups. Pupils who received no feedback, on the other hand, tended to attribute success to the examiner's mood, $F(2, 258) = 17.66, p < .001$, and to neatness, $F(2, 258) = 43.06, p < .001$, significantly more than did pupils who received grades or written comments. Although these results are quite complex, they do in general suggest that pupils who received written comments mostly tended to attribute success to internal, motivational factors such as effort and interest. Subjects who received no feedback, although scoring high on attribution to ability, also tended more than those in other groups to attribute success to external or "irrelevant" factors. Subjects who received grades attributed success to internal factors less often than did those who received comments, and attributed success to external factors less often than did those who received no feedback.

Preference for mode of evaluation. The percentage of subjects in each group expressing preference for each mode of evaluation are presented in Table 4. Most pupils chose written comments, but the intergroup patterns are also striking. The comments group showed the strongest preference for its own mode of evaluation, although almost as many pupils in the grades group also chose written comments. Subjects who received no feedback were equally divided in their preferences for written comments and numerical grades; none of them chose to receive no feedback. The association

Table 4
Percentages of Subjects Expressing Preference for Each Mode of Evaluation, by Experiment Group

Mode of evaluation	Comments group	Grades group	No feedback group	All groups
Written comments	86.3	78.9	51.8	73
Numerical grades	10.2	21.0	48.2	25
No feedback	3.4	0.0	0.0	1

between experimental group and preference for evaluation mode was highly significant, $\chi^2(4, N = 261) = 43.66, p < .001$.

Discussion

The present study was designed to test the effects of different feedback conditions on continuing motivation (Maehr, 1976) — that is, on later performance of and attitudes toward tasks originally perceived as highly interesting. On the basis of theoretical considerations we hypothesized that informative comments would serve to maintain or even to enhance intrinsic motivation by answering needs for mastery and self-evaluation. Even though the findings of much previous research, which compared motivation after a single trial under extrinsic and no-incentive conditions, implied that if "left alone," intrinsic motivation for an initially interesting task would remain stable and high, our conception suggested that, in fact, repeated nonreceipt of feedback would undermine interest. We also predicted that numerical grades would foster extrinsic motivation at the expense of intrinsic motivation. The results showed clearly differential effects of these kinds of feedback on performance. The comments group scored high on both tasks at posttest (and higher than they had at pretest), whereas the no-feedback group scored low (and much lower than at pretest) on both. The grades group scored high only on the quantitative task and scored low (and lower than at pretest) on the divergent thinking measures, which have been found to be adversely affected by extrinsic incentives (cf. Kruglanski, Friedman, & Zeevi, 1971).

Although the motivational framework suggested above seems to provide an inclusive explanation for these differences in performance, it can be argued, instead, that they derived primarily from the effects of feedback in increasing the salience of success in the task and, in the case of comments, also of the criteria relevant to such success. It is indeed probable that some such effects occurred, and the particular influences of specific feedback on subsequent performance seem worthy of further attention. However, several considerations seem to reduce the power of this argument to explain the results of the study as a whole. First, the comments did not include any information additional to that repeatedly given to all groups before each session. In addition, the comments group also scored highest on elaboration, an aspect of Task B not mentioned in either the instructions or the comments. Second, feedback-induced changes in the salience of successful performance do not explain why the group that received grades, presumably

Table 3
Percentages of Subjects Attributing Effort to Each Factor in Each Experimental Group

Factor	Comments group	Grades group	No feedback group
Interest	70.5	34.4	43.4
Felt it important to succeed	17.0	34.4	40.0
Thought scores would influence school grades	3.4	4.4	7.2
Desire to avoid poor achievement	9.0	26.7	9.6

a potent cue as to the importance of success, scored high only on the quantitative task. Third, since the experimenter was not known to the students, who were also specifically told that they would *not* be evaluated on their Session 3 performance, it is doubtful whether the differences obtained were due in any large part to treatment-induced differences in social desirability.

These considerations lead us to an explanation in terms of intrinsic motivation. This interpretation is based on a well-developed theoretical framework, presented in brief in the introduction. It provides an inclusive explanation for the overall pattern of the performance data and receives consistent support from the results of the questionnaire. As hypothesized, subjects who received written comments expressed greater interest in the tasks than did those in the other two groups, especially for the question requiring greatest commitment — the number of extra tasks chosen. The patterns of attribution do not in themselves provide a measure of intrinsic motivation, but they do help clarify the motivational climates induced by each feedback condition. Although the attribution patterns of all groups reflected the predominantly internal orientation usually found in the middle class (Rotter, 1966), group differences within this general pattern were suggestive. To a greater extent than the grades and no-feedback groups, the comments group saw interest as determining both effort and success and viewed arbitrary or other-controlled factors (such as the examiner's mood) as unimportant. This pattern is consistent with the hypothesis that task-related comments would foster a climate characterized by high interest and personal causation. Moreover, it should be remembered that the feedback consisted not only of a "positive" comment relating to some aspect of the task that was performed well but also of a "negative" one relating to some aspect performed less adequately. Thus our findings can less easily be attributed to the effects of self-enhancement alone than can those of other studies that provided only positive information.

The motivation of students who received grades is also clarified by their patterns of attribution of effort. Thus, over 50% attributed it to the desire to succeed or to avoid failure, the two components of achievement motivation (Atkinson, 1957). Need achievement may well be the motive routinely dominant in school settings, unless interest is exceptionally high. In addition, in many students the use of grades seems to activate the anxiety behind the motive to avoid failure (Atkinson & Litwin, 1960). Indeed, in the grades group, 26.7% attributed effort to the desire to avoid poor scores, and 34.4% attributed it to the importance of success, whereas among those who received no feedback only 7.2% attributed effort to the desire to avoid poor scores and 40% attributed it to the "positive" aspect of need achievement, the desire to succeed. In the no-feedback group, neither interest nor achievement motivation were strong enough to maintain, let alone enhance, original levels of performance in either task. In addition, repeated non-receipt of feedback seemed to induce confusion as to the source of motivation, which was reflected, for example, in the tendency of some of the subjects to misattribute success to irrelevant factors.

These results seem to have important implications both for the study of intrinsic motivation and for its enhancement in the

schools. Several authors have commented on the insufficient attention paid to date to the study of the maintenance and continuance of, as opposed to the engagement or reengagement in, intrinsically motivated behavior (Condry, 1977; Machr, 1976). Our findings suggest that intrinsic motivation is not merely a function of collative stimulus properties (Berlyne, 1960) but depends on the dynamic interaction between the stimulus and the individual — that is, on the degree to which a task continues to be perceived as challenging and as providing satisfying increments in one's knowledge about one's competence. Thus the availability and nature of feedback will importantly influence whether interest is maintained or undermined.

Our results suggest, as some critics argue (Holt, 1964; Silberman, 1970), that the information routinely given in schools — that is, grades — may encourage an emphasis on quantitative aspects of learning, depress creativity, foster fear of failure, and undermine interest. They also suggest that no such negative results ensue from the use of task-related individualized comments. The differential effects of these modes of evaluation on performance and motivation seem to derive not only from the salient control aspect of grades discussed earlier but also from differences in the information they provide: Normative grades provide information about proficiency relative to others; they do not provide clear standards for self-evaluation or for constructive attribution (Nisan, 1981). Our finding that 50% of the pupils who received no feedback would rather have received a grade may seem compatible with the arguments of many teachers that pupils themselves want grades. However, it should be remembered that 78.9% of the pupils who actually received grades would have preferred written comments, and 86.3% of those who received comments were satisfied with this mode of evaluation. Thus most pupils seem to prefer normative information to no information; however, they also prefer over grades the kind of constructive, specific information about competence provided by the written comments.

Finally, we should note some significant limitations in deriving guidelines for evaluation applicable in school settings from our results. First, every pupil could succeed to some extent on the experimental tasks, and none experienced absolute failure. However, not all school tasks are of this kind, and one can envisage situations in which evaluation will be primarily positive or negative, rather than balanced. Second, the tasks employed in this study were "interesting" to start with, whereas many of those learned in school are not. Finally, although our three-trial experimental paradigm is an improvement on most existing studies, it still far from approximates the situation in the school. It would certainly be valuable to attempt to study the effects of different modes of evaluation in a more ecologically representative setting.

References

- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*, 359–372.
- Atkinson, J. W., & Litwin, G. M. (1960). Achievement motivation and test anxiety conceived as the motive to approach success and the motive to avoid failure. *Journal of Abnormal and Social Psychology*, *60*, 52–63.

- Berlyne, D. E. (1960). *Conflict, arousal and curiosity*. New York: McGraw-Hill.
- Boggiano, A. K., & Ruble, D. N. (1979). Competence and the over justification effect: A developmental study. *Journal of Personality and Social Psychology*, 37, 1462-1468.
- Condry, J. (1977). Enemies of exploration. *Journal of Personality and Social Psychology*, 35, 459-477.
- DeCharmes, R. (1968). *Personal causation*. New York: Academic Press.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105-115.
- Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum Press.
- Deci, E. L., Cascio, W. F., & Krusell, J. (1973, May). *Sex differences, positive feedback and intrinsic motivation*. Paper presented at the meeting of the Eastern Psychological Association, Washington, DC.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Harter, S. (1978). Pleasure derived from challenge and the effects of receiving grades on children's difficulty level choices. *Child Development*, 49, 788-799.
- Harter, S. (1981). A model of mastery motivation in children: Individual differences and developmental change. In W. A. Collins (Ed.), *Minnesota Symposia on Child Psychology: Aspects of the development of competence*, (Vol. 24, pp. 215-255). Hillsdale, NJ: Erlbaum.
- Holt, J. (1964). *How children fail*. New York: Dell.
- Kruglanski, A. W., Friedman, I., & Zeevi, G. (1971). The effects of extrinsic incentives on some qualitative aspects of task performance. *Journal of Personality*, 39, 606-617.
- Maehr, M. L. (1976). Continuing motivation: An analysis of a seldom considered educational outcome. *Review of Educational Research*, 46, 443-462.
- McGraw, K. O. (1978). The detrimental effects of reward on performance: A literature review and a prediction model. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of rewards* (pp. 33-60). Hillsdale, NJ: Erlbaum.
- Nisan, M. (1981). Motivational aspects of evaluation in school. In I. A. Lewy & D. Nevo (Eds.), *Evaluation roles in education* (pp. 133-137). London: Gordon and Breach.
- Pittman, T. S., Davey, M. E., Alafat, K. A., Vetherill, K. V., & Kramer, N. A. (1980). Informational and controlling verbal rewards, levels of surveillance and intrinsic motivation. *Personality and Social Psychology Bulletin*, 6, 228-233.
- Rotter, J. B. (1966). Generalized expectancies for internal vs. external control of reinforcement. *Psychological Monographs*, 30, 1-26.
- Silberman, C. (1970). *Crisis in the classroom*. New York: Random House.
- Suls, J. M., & Miller, R. L. (Eds.). (1977). *Social comparison processes: Theoretical and empirical perspectives*. New York: Hemisphere.
- Torrance, A. E., & Templeton, O. E. (1963). *Minnesota Tests for Creative Thinking*. Minneapolis: University of Minnesota Press.

Received May 4, 1984

Revision received March 19, 1985 ■

Low Publication Prices for APA Members and Affiliates

Keeping You Up-to-Date

All APA members (Fellows, Members, and Associates) receive — as part of their annual dues — subscriptions to the *American Psychologist*, the *APA Monitor*, and *Psychology Today*.

High School Teacher and Student Affiliates receive subscriptions to the *APA Monitor* and *Psychology Today*, and they can subscribe to the *American Psychologist* at a significantly reduced rate.

In addition, all members and affiliates are eligible for savings of up to 50% on other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the British Psychological Society, the American Sociological Association, and Human Sciences Press).

Essential Resources

APA members and affiliates receive **special rates** for purchases of APA books, including the *Publication Manual of the APA*, the *Master Lectures*, and *APA's Guide to Research Support*.

Other Benefits of Membership

Membership in APA also provides eligibility for low-cost insurance plans covering life; medical and income protection; hospital indemnity; accident and travel; Keogh retirement; office overhead; and student/school, professional, and liability.

For more information, write to American Psychological Association,
Membership Records, 1200 Seventeenth Street NW, Washington, DC 20036