Journal of Educational Psychology
1987, Vol. 79, No. 2, 137-146

0022-0663/87/$00.75

Differential Effects of Teacher Comments

Günter Krampen
University of Trier, FRG

An analysis of the effects of three kinds of teacher comment (social-comparison, subject-matter, and intraindividually oriented) accompanying grades in mathematics is presented. Subjects were 385 students from Grades 6 through 10 who were randomly assigned to one of three groups, each receiving one of the three types of comment, or to a control group. During the first half of the school year, teachers wrote comments concerning students' performance on mathematics examinations; during the second half, comments were not made so as to test the duration of effects resulting from the earlier comments. Students were observed at the beginning of the experiment, at the end of the treatment period, after the no-comment phase, and every time they received a grade in mathematics. Dependent variables (besides later performance in mathematics) were cognitive-motivational variables and some school-related personality characteristics, which were selected with reference to a differentiated expectancy-value theory of achievement motivation. In analyses of variance, an interaction between type of comment and the student's prior performance level was revealed for almost all dependent variables. Effects of comments persisted after the experimental period for school-related personality variables only (some aspects of scope-specific locus of control and test anxiety). Implications for teaching strategies as well as for developmental and educational psychology are suggested.

The results of the effects of written teacher comments about exam grades on student performance are inconsistent. Page (1958) reported positive effects (see, e.g., Mathis, Cotton, & Sechrest, 1970; Schmidt, 1978). Stewart and White (1976), however, surveyed 12 attempts to replicate Page's findings and found little agreement between Page's major finding and the replication studies. They concluded, somewhat cautiously, that (a) "there is no strong evidence to suggest that any type of comment retains its effectiveness over an extended period of time or if administered on more than one occasion" (Stewart & White, 1976, p. 498); (b) there is no consistent evidence for comment effect at the elementary or secondary school level, although slight evidence is found at the college level; and (c) where comments were effective, "they were encouraging and personalized in nature, rather than simple, standard statements" (Stewart & White, 1976, p. 498). Recently, Elawar and Corno (1985) found strong main effects on achievement and attitude for individually oriented teacher comments on student homework (not exams).

Most of the research has no theoretical base and is unrealistic in the sense of lacking external validity. These features of the research may be the reason for the uneven findings. In this study, I developed a theoretical perspective for research on teacher comments and student performance and then tested it in an empirical study.

Dependent Variables and Duration of Investigation

Page (1958) and most of those who followed restricted themselves to the measurement of grades or other achievement scores as dependent variables. When short time periods are studied (3- to 10-week periods), these dependent variables are unlikely to change very much. Studies should cover a longer time period so that it is possible to study the cumulative and long-term effects of teacher comments. They should also include time periods during which comments are not made so that the duration of effects resulting from earlier comments can be examined.

The selection of dependent variables is also critical. The effects of comments are probably mediated by cognitive-motivational variables. Cognitive-motivation theory provides a conceptual framework for such variables (see, e.g., Atkinson & Birch, 1970; Heckhausen, 1977). Situation- and action-specific variables such as subjective expectation of improvement and subjective valuation of improvement (valence) have proved significant as predictors of academic achievement (e.g., Batlis, 1978; Grobe & Hofer, 1983; Henson, 1976; Krampen, 1979; Krampen & Lehmann, 1981; Misanchuk, 1977; Mitchell & Pollard, 1973).

On the basis of the expectancy-value theory of motivation, it makes sense to extend the set of dependent variables in studies analyzing the effects of teacher comments beyond achievement measures and grades (later performance) to situation- and action-specific variables as well as to generalizations from them (like locus of control of reinforcement and test anxiety; see Krampen, 1984; Pekrun, 1983; Rotter, 1966). Although these variables have a higher potential for both inter- and intraindividual variability (in comparison with pure achievement measures and grades), some groups of students must be preselected for some analyses. For example, pupils with very high grades must be eliminated from analyses of valence of improvement because it is not possible for them to improve their grades.

Specification of Type of Comment

My second criticism of most earlier research in this field is the inadequate specification of the independent variable. With

The research was supported by Grant Kr 788/1-1/2 of the Deutsche Forschungsgemeinschaft.

Correspondence concerning this article should be addressed to Günter Krampen, Department of Psychology, University of Trier, PO 3825, D-5500 Trier, Federal Republic of Germany.

reference to different models of grading (e.g., Klauer, 1982) and of reference-norm orientation (e.g., Rheinberg, 1980), the following three types of teacher comments can be distinguished: (a) socially oriented comments, focused on interindividual comparisons of achievement (paraphrasing letter grades; e.g., "In comparison with the other pupils, your achievement is very bad"); (b) subject-matter-oriented comments, based on absolute standards (achievement criteria; e.g., "You know how to solve the task—the formula is okay—but your computation is wrong in this instance . . ."); and (c) individually oriented comments, based on intraindividual comparisons of achievement (considering each student's learning history and achievement record; e.g., "In comparison with your prior performance, you have improved a lot . . ."). In most of the studies, the three types of comments have been mixed.

A final point is the importance of investigating interactions as well as main effects. In the present case, the issue is the differential effect of written teacher comments, depending on the student's prior performance level, a form of aptitude–treatment interaction. The common application of unspecific comments, without regard to the individual level of prior performance, is supposed to give away valuable information about the differential efficacy of commenting. One of the goals in what follows, therefore, is an answer to the question of which type of written teacher comment resulting in interaction with a student's prior performance level has the most positive effects?

Expectancy Effects

Students have generally not been informed that they were participating in an investigation. Expectancy effects on the teachers' part have been ignored by and large. Whenever a teacher is supposed to react in a specific way to some students in a class and not to others (Fittkau & Langer, 1974; Page, 1958; Rheinberg, Kühmel, & Duscha, 1979), expectancy effects may occur. This is also true if a teacher has to make different types of comments in two or more classes. The traditional solution to this problem, the double-blind technique, is out of the question when individually oriented comments have to be made as part of the experiment.

One solution might be to have each teacher implement just one experimental condition, thus avoiding differential effects for different experimental groups. However, experimental groups, which ideally should consist of several classes in each condition, have to go in tandem with their teachers. The unit of randomization is the class, a preexisting group; both the experimental group and the control group consist of several undivided classes. The assignment of classes and their teachers to conditions must be done randomly. The unit for data analyses is then the class (see Cook & Campbell, 1979). When the comparability of individuals from different classes can be assured with reference to those variables that may offer alternative explanations for experimental effects, the unit is the individual. To exclude student expectancy effects, researchers should integrate the investigation into regular classroom procedures for the entire school year.

I selected mathematics education for this study because (a) in the schools from which the participating classes were drawn, mathematics teachers had not been issued any directives regarding comments on grades; (b) grades in mathematics are somewhat more objective than grades in other school subjects; and (c) the math curriculum is more standardized from class to class and from school to school.

The empirical study had two purposes. The first was to test the hypothesis of different effects of the three types of written teacher comments about grades on later performance, including cognitive-motivational variables and attitude and personality variables. The second was to test both main effects and interactions. Thus, I reached a two-step specification and differentiation of the traditional and molar examination of relations between teacher comments and student variables. To examine the duration of effects, I did this in a longitudinal design including periods with and without teacher comments.

Method

Experimental Design and Subjects

In cooperation with the education department of the Trier County government, I contacted all *Realschulen* (comparable with junior high schools) of Trier County and township in the summer of 1982 and asked them to participate in the investigation. The schools' mathematics teachers were asked to send back a short questionnaire concerning grade levels taught and grading procedures. From the responses, I selected 13 classes. None of the teachers were accustomed to adding written comments when grading math exams. Thirteen teachers and 385 students participated in the experiment (190 girls and 195 boys; mean age = 13.9 years, $SD = 1.40$; one 6th-grade class, two 7th-grade classes, two 8th-grade classes, four 9th-grade classes, and four 10th grade classes). Class size ranged from 24 to 34 students. All the students involved agreed to participate in the data collection and were briefly informed about the organization of the study—without, however, any reference to the experiment.

Controlling for grade level, I randomly assigned each of the 13 classes and their teachers to the following four groups: Experimental Group 1, socially oriented comments on all math exams during the first semester of the school year (Grades 6, 9, and 10; $n = 87$); Experimental Group 2, subject-matter-oriented comments on all math exams during the first semester (Grades 8, 9, and 10; $n = 116$); Experimental Group 3, individually oriented comments on all math exams during the first semester (Grades 7, 8, 9, and 10; $n = 92$); and Experimental Group 4, the control group, no written comments on math exams (Grades 7, 9, and 10; $n = 90$).

The math teachers received short individual instructions about the type of comments to be made. (For definitions and examples of the comment types, see the Specification of Type of Comment section.) Each teacher had to use just one type of comment (including no comment). Instruction manuals explained the requested procedure in more detail.

Each time exams were returned to students, research assistants randomly tested teachers' comments: They checked fidelity of treatment implementation for 6 to 10 students per class. The results of those procedures confirmed a high level of fidelity of treatment implementation; 86% of the checked comments were totally right, and 14% were only somewhat wrong, that is, included other than the intended aspects of feedback. Each teacher was informed only about the type of comment he or she was to use. After semester grades were

given in February 1983, all comments on exams ceased, but data gathering continued. The teachers were remunerated for their extra work (writing comments) from a grant made by the Deutsche Forschungsgemeinschaft.

Besides an extensive pretest at the beginning of the 1982–1983 school year in September (t_1), the students took an extensive first posttest immediately after the semester grade reports were handed out in February (t_5) and an extensive second posttest after final grade reports were handed out in July (t_9). In addition, the grades of all written math exminations during the school year were recorded (t_2 through t_4 and t_6 through t_8). All data were gathered anonymously by experimenters, who were careful to see that teachers did not influence the students while they were filling out the questionnaires. Questionnaires were assigned to students by using a stable person code.

Methods of Data Collection

At the pretest, students completed the following four questionnaires:

1. The Learning Situation Test (LST; Kahl, 1977) was used to assess subjective perceptions of classroom atmosphere. The LST consists of 65 items marking four factor analytically derived scales measuring (a) perceived cohesion in the class, (b) identification with instructions, (c) perceived demands for performance, and (d) perceived degree of competition and order. In the present sample, split-half reliability of the LST subscales exceeded .78.
2. A special questionnaire was used to assess level of performance in mathematics as well as situation-specific cognitive-motivational variables; among others, the following variables were measured: mathematics grade for the last term; motivation for improvement in mathematics (6-point scale); subjective valuation, or valence, and subjective expectancy of improvement (two 7-point bipolar rating scales); and perceived overload in math lessons (yes or no). Checks of the internal consistency confirmed the reliability of this data in the present sample.
3. The IPC-PL Questionnaire was used to assess three aspects of locus of control for problem-solving behavior (internality, expectancy for control by powerful others, and belief in chance); each of these three scales consists of eight items. Split-half reliabilities ($r_{tt} \geq .69$) confirmed the fitness of the IPC-PL in the present sample for group comparisons.
4. The Anxiety Questionnaire for Students (AFS; Wieczerkowski, Nickel, Janowski, Fittkau, & Rauer, 1975) was used to assess (a) manifest anxiety, (b) test anxiety, (c) dislike or aversion toward school, and (d) social desirability. Split-half reliability in the present sample exceeded .72.

During the two posttests, which terminated the first and second experimental phases, the instruments used in the pretest, except for the LST, were applied.

Methods of Data Analysis

Data analysis was done by multivariate and univariate analyses of variance and contrast tests (Duncan's multiple-range tests). The significance level was fixed at .05. So that the importance of the results may be judged, estimates of effect size—measures of practical significance—are given, in addition to measures of statistical significance (Cohen, 1977).

Data were analyzed in a two-step procedure. First, to exclude possible alternative explanations arising from a priori differences between classes and students, I tested the comparability of the experimental groups. Class effects were treated the same way: Analyses of variance (ANOVAS) tested whether the variance between classes was significantly greater than that within classes. If it was not, the individual was then chosen as the unit of analysis (cf. Cook & Campbell, 1979). The results of these procedures were obtained by more conservative computations that had class means as the unit of analysis. Students for whom data were missing were omitted only from those analyses that required the missing variables. Data analysis was done separately for (a) grades and cognitive-motivational variables and (b) attitude and personality variables. Combined analyses made no sense because these sets of variables differed in their situation specificity.

Results

Comparability of Experimental Groups

Means, standard deviations, and intercorrelations of the variables in the two sets for all pre- and postmeasures are summarized for the entire sample in Tables 1 through 4. The correlation patterns agreed with the expectations and previous findings (see Batlis, 1978; Grobe & Hofer, 1983; Henson, 1976; Krampen, 1979; Krampen & Lehmann, 1981; Pekrun, 1983).

To check the comparability of the four experimental groups, I computed unifactorial multivariate analyses of variance (MANOVAS) for the LST, IPC-PL, and AFS scales $F(30, 991) = 0.470$, $MS_e = 3{,}227.45$, and for cognitive-motivational variables and last grade in mathematics, $F(27, 1066) = 0.214$, $MS_e = 2{,}511.29$, as measured at the pretest. Univariate ANOVAS and Duncan's multiple-range tests also failed to show any significant differences among the four groups; Bartlett-Box tests gave no evidence of heterogeneity of variances. Thus, at the beginning of the experiment, the four experimental groups were comparable in achievement-related variables, cognitive-motivational variables, various aspects of class atmosphere, and selected personality characteristics (aspects of locus of control and of anxiety). This is shown in detail in Tables 1 and 2.

Class differences could also be excluded for the pretest data. MANOVAS computed with the variable of class membership failed to show for cognitive-motivational variables and grade, $F(108, 2706) = 0.893$, $MS_e = 237.55$, as well as for personality and attitude variables, $F(120, 2832) = 0.937$, $MS_e = 265.13$, that the variance among the 13 classes was significantly greater than the variance within them.

Effects on Cognitive-Motivational Variables and Reported Grades

For the t_5 data (first posttest after the commenting condition), ANOVAS yielded significant main effects on cognitive-motivational variables and on reported grades (see Table 5). Thus, the first hypothesis was confirmed: The different experimental conditions (types of comments) resulted in significant differences in grade and in expectancy of and motivation for improvement measured at the first posttest.

These main effects had to be differentiated because significant interactions were found, too, which confirmed the hypothesis of aptitude–treatment interaction. This is shown in detail in Table 5. Commenting did not affect subjective

Table 1
Means and Standard Deviations of Cognitive-Motivational Variables and Reported Grades in the Experimental Groups ($225 \leqq N \leqq 385$)

Variable	Pretest (t_1)		First posttest (t_5)		Second posttest (t_9)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Reported grade						
Group 1	3.4	0.78	3.6	0.81	3.5	0.79
Group 2	3.5	0.97	3.2	0.96	3.4	1.04
Group 3	3.4	0.69	3.0	0.74	3.3	0.72
Group 4	3.3	0.81	3.4	0.87	3.4	0.85
Valence of improvement[a]						
Group 1	2.3	0.93	2.1	1.07	—	—
Group 2	2.2	1.01	1.9	1.21	—	—
Group 3	2.3	1.07	2.1	1.28	—	—
Group 4	2.1	0.96	2.2	1.07	—	—
Expectancy of improvement[a]						
Group 1	1.3	1.19	0.5	1.31	—	—
Group 2	0.9	1.31	1.1	2.71	—	—
Group 3	1.2	1.21	2.0	2.01	—	—
Group 4	1.1	1.25	0.9	1.25	—	—
Motivation for improvement[a]						
Group 1	3.6	1.39	3.1	1.60	—	—
Group 2	3.7	1.55	4.0	1.60	—	—
Group 3	3.8	1.55	4.3	1.77	—	—
Group 4	3.7	1.58	3.6	1.65	—	—

Note. Group 1 = socially oriented comments; Group 2 = subject-matter-oriented comments; Group 3 = individually oriented comments; Group 4 = control.
[a] Not measured at t_9 because many students finished school.

valence of improvement in mathematics, but it had—differentially for students with different performance levels in mathematics—a significant effect on reported grades and on expectancy of and motivation for improvement. Estimates of effect size in Table 5 demonstrate that the impact of experimental conditions was higher on expectancy of improvement than on motivation for improvement. This corresponds to expectancy-value theories of motivation, in which a multiplicative relation between valences and expectancies in the prediction of motivational variables is postulated. According to Cohen's (1977) criteria, the observed effect sizes of commenting are small to medium.

The interaction between type of comment (experimental condition) and level of performance in math at t_1 is shown in Figure 1 exemplarily for the cognitive-motivational variable of expectancy of improvement in mathematics. This elucidates the full confirmation of the interaction hypothesis: Socially oriented teacher comments resulted in very low expectancies of improvement for low-performing students and in high expectancies for students with satisfactory performance. Thus, this type of comment further accentuated existing differences between performance groups. Subject-matter-oriented comments as well as individually oriented comments showed positive effects for all performance levels (compared with the control group); individually oriented comments had a significantly (Duncan's multiple-range test, $p < .05$) stronger impact on expectancy of improvement than did subject-matter-oriented comments.

For the variable of grade from the term grade reports (t_5), Figure 2 shows the interaction of the variables of type of comment/experimental condition and level of performance at pretest (t_1). In substance, the findings were very similar to those for the situation-specific cognitive-motivational variables. Whereas socially oriented comments further accentuated differences between performance groups, all students improved to some degree as a result of individually oriented and subject-matter-oriented comments. Again, individually oriented comments were significantly (Duncan's multiple-range test, $p < .05$) superior to subject-matter-oriented comments, especially for low-performing students. The estimates for effect size were again either small or medium, except for the variable of level of performance (t_1). In this case, the magnitude corresponded to results of correlation studies, which demonstrate again and again that the best predictors of grades are grades received earlier (e.g., Schwarzer, 1979).

No comments were made at the time of the second posttest (t_9), half a year after the first posttest. By then, all effects of comment on grades had disappeared (see Tables 1 and 5).

The interaction hypothesis was thus confirmed by the results for the cognitive-motivational variables and semester grades in mathematics. The hypothesis was valid, however, only at the first posttest, which was immediately after the first experimental stage in which intervention (comment) occurred. No long-term effects of teacher comments were observed for those variables.

Effects on Personality and Attitude Variables

The three scales of the IPC-PL Questionnaire measuring locus of control for intellectual problem-solving behavior and

Table 2
Means and Standard Deviations of Attitude and Personality Variables in the Experimental Groups (225 ≦ N ≦ 385)

Variable	Pretest (t_1)		First posttest (t_5)		Second posttest (t_9)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Internality						
Group 1	34.4	5.61	34.8	6.41	34.8	6.32
Group 2	34.5	5.24	34.6	6.22	34.4	5.89
Group 3	34.6	4.79	37.2	4.83	36.7	4.91
Group 4	34.8	5.41	34.0	6.22	34.2	5.89
Powerful others externality						
Group 1	25.1	5.95	28.9	6.21	27.3	6.05
Group 2	23.9	5.63	21.0	6.04	24.0	6.00
Group 3	23.3	5.47	21.2	6.09	22.8	5.87
Group 4	25.1	5.45	24.7	6.10	25.3	6.13
Chance control						
Group 1	26.8	5.24	28.7	5.88	27.5	6.03
Group 2	25.4	5.31	23.2	6.14	24.9	5.81
Group 3	25.8	5.78	22.4	5.61	23.1	6.03
Group 4	27.2	5.23	26.9	5.73	27.3	5.61
Attitude toward school						
Group 1	15.3	2.56	16.9	2.37	16.4	2.41
Group 2	15.5	2.82	15.6	2.91	15.6	2.87
Group 3	15.1	2.49	14.8	2.60	15.0	2.58
Group 4	15.5	2.45	15.4	2.71	15.6	2.62
Test anxiety						
Group 1	21.8	4.29	22.0	4.37	21.9	4.40
Group 2	21.5	4.34	22.4	4.63	22.3	4.79
Group 3	21.2	3.71	18.3	4.12	19.9	4.03
Group 4	21.3	4.02	21.8	4.62	21.9	4.70
Manifest anxiety						
Group 1	23.4	3.79	23.7	4.10	23.6	4.04
Group 2	23.6	3.85	23.9	4.33	23.9	4.09
Group 3	23.4	3.28	21.4	3.94	22.9	3.98
Group 4	22.9	3.67	23.6	4.22	23.4	4.11

Note. Group 1 = socially oriented comments; Group 2 = subject-matter-oriented comments; Group 3 = individually oriented comments; Group 4 = control.

the three scales of the AFS—the ones measuring manifest anxiety, test anxiety, and negative attitude toward school—were likewise used as dependent variables in ANOVAS. Whereas there were no group differences in these variables at pretest, data obtained at the first posttest revealed significant main effects for performance level at t_1 and for experimental condition and some significant interactions (see Tables 2 and 6). The data obtained at the second posttest yielded weaker but still significant main effects yet no significant interaction of the two variables.

Table 3
Intercorrelations Between Pretest and Posttest Measures of Cognitive-Motivational Variables and Reported Grades (225 ≦ N ≦ 385)

Measure	2	3	4	5	6	7	8	9
Pretest (t_1)								
1. Reported grade	−.40	.21	.54	.68	−.18	.20	.33	.48
2. Valence of improvement		.30	.48	−.22	.43	.18	.31	−.23
3. Expectancy of improvement			.43	.18	.12	.54	.31	.17
4. Motivation for improvement				.61	.33	.37	.60	.48
First posttest (t_5)								
5. Reported grade					−.31	.25	.48	.65
6. Valence of improvement						.34	.52	−.25
7. Expectancy of improvement							.48	.38
8. Motivation for improvement								.59
Second posttest (t_9)								
9. Reported grade								

Note. $r \geqq |.12|$, $p < .05$.

Table 4
Intercorrelations Between Pretest and Posttest Attitude and Personality Measures (225 ≧ N ≧ 385)

Measure	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Pretest (t_1)																	
1. I	−.35	−.23	−.23	−.23	−.13	.68	−.21	−.05	−.01	−.18	−.09	.59	−.19	−.22	−.11	−.20	−.11
2. P		.57	.30	.38	.35	−.29	.77	.29	.24	.36	.04	−.18	.51	.21	.17	.20	.09
3. C			.23	.30	.34	−.19	.37	.81	.17	.22	.24	−.17	.32	.78	.19	.23	.14
4. ATS				.34	.23	−.09	.24	.25	.59	.17	.11	−.11	.19	.17	.48	.10	.13
5. TA					.68	−.20	.17	.18	.08	.61	.39	−.12	.09	.09	.05	.41	.29
6. MA						−.11	.20	.25	.19	.40	.73	−.13	.07	.21	.11	.35	.68
First posttest (t_5)																	
7. I							−.28	−.29	−.19	−.30	−.15	.64	−.19	−.11	.05	−.12	−.11
8. P								.51	.34	.32	.39	−.21	.69	.20	.19	.29	.09
9. C									.26	.39	.40	−.20	.40	.77	.20	.20	.20
10. ATS										.28	.14	−.08	.19	.19	.54	.15	.12
11. TA											.59	−.14	.11	.15	.10	.62	.40
12. MA												−.07	.18	.27	.18	.43	.65
Second posttest (t_9)																	
13. I													−.19	−.21	−.23	−.17	−.09
14. P														.40	.30	.31	.35
15. C															.28	.40	.43
16. ATS																.24	.08
17. TA																	.65
18. MA																	

Note. I = internality; P = powerful others externality; C = chance control; ATS = attitude toward school; TA = test anxiety; MA = manifest anxiety. $r \geqq |.12|$, $p < .05$.

Table 6 shows the results of the ANOVAS for both posttests. Data obtained at t_5 (first posttest) showed the predicted main and interaction effects for internality, powerful others externality, negative attitude toward school, and test anxiety. The scales for chance control (fatalistic locus of control orientation for problem solving) and for manifest anxiety, however, revealed only main effects for experimental condition (type of teacher comment), thus confirming the main-effect hypothesis, not the aptitude-treatment-interaction hypothesis.

Four of the six scope-specific personality and attitude variables—internality in locus of control, powerful others control, negative attitude toward school, and test anxiety—showed long-term effects, as demonstrated by the significant main effects of experimental conditions for the data obtained at t_9, about half a year after the last experimental treatment. In particular, socially oriented teacher comments had especially negative effects on powerful others control orientations for intellectual problem solving and on negative attitude toward school. With individually oriented comments, internality in locus of control for problem solving increased, and chance control for problem-solving behavior as well as test anxiety decreased.

Effects on Grades on Mathematics Examinations

With respect to the data concerning the math exams of the students during the school year, I present the results for the examination grades in Table 7. At t_2, there were no significant intergroup differences in grades on mathematics exams. This examination was the first to be commented on in the experimental groups; thus, no effects could be expected. By t_3 (second exam of the comment phase), significant group differences appeared. As confirmed by t tests for dependent samples, subject-matter-oriented comments resulted in deterioration of the average grade; by contrast, individually oriented comments resulted in improvement. At t_4, further improvement for Group 3 (individually oriented comment) and substantial improvement for Group 2 (subject-matter-oriented comment), which resulted in significant intergroup differences at t_4, could be observed ($p < .05$). These changes and the intergroup differences remained constant at t_5 (term grade report) and t_6 (first exam after comment stopped). They receded, however, at t_7 and t_8. This corresponded to the findings on term grades.

Group 1 (socially oriented comment), however, did not show statistically significant changes in grades as a whole. This can be explained by the interaction between performance level, which has so far in this section been ignored, and type of comment. Further deterioration of low-performing students and further improvement (to the extent possible) of high-performing students neutralized each other in group means. Thus, the cumulative effect of teacher comment was confirmed. A truly cumulative effect, however, could be observed only for individually oriented comments. Subject-matter-oriented teacher comments led first to a decline in performance, which was, of course, offset pretty soon. The importance of the interaction between performance level and type of comment again was clearly shown for the socially oriented type of written teacher comment.

Table 5
Analyses of Variance (ANOVAs) on Posttest Measures of Cognitive-Motivational Variables and Reported Grades Using Treatment and Prior Performance Level as Variables

Dependent variable and posttest	ANOVA effect: Treatment F	Treatment ω^2	Performance level F	Performance level ω^2	Interaction F	Interaction ω^2	MS_e
Reported grade							
t_5	19.77*	0.081	60.04*	0.338	2.22*	0.021	0.53
t_9	1.57		29.39*	0.241	0.22		1.72
Valence of improvement							
t_5	1.47		13.88*	0.211	0.36		1.62
Expectancy of improvement							
t_5	11.77*	0.123	6.54*	0.061	4.77*	0.053	2.15
Motivation for improvement							
t_5	10.57*	0.104	6.42*	0.053	3.61*	0.042	1.87

Note. The ω^2 is given only if statistical significance was reached.
* $p < .05$.

Analyses on Class Level

The results, based on the individual student as the unit of statistical analysis, are protected against the possible criticism of crude randomization by further computations based on class means as the unit of analysis. An ANOVA was computed for the means of the term grades (t_5) in the 13 participating classes. Again, independent variables were prior performance

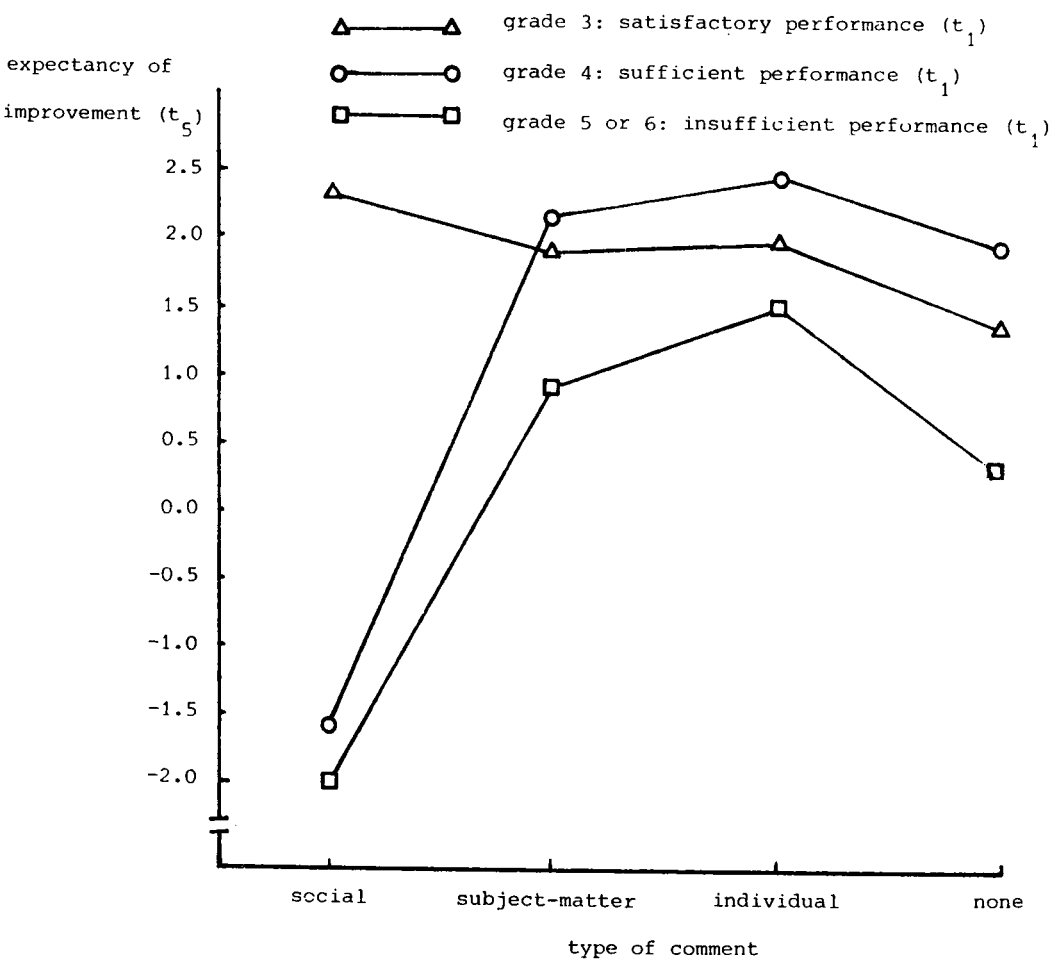

Figure 1. Type of Comment × Prior Performance Level interaction for expectancy of improvement at t_5.

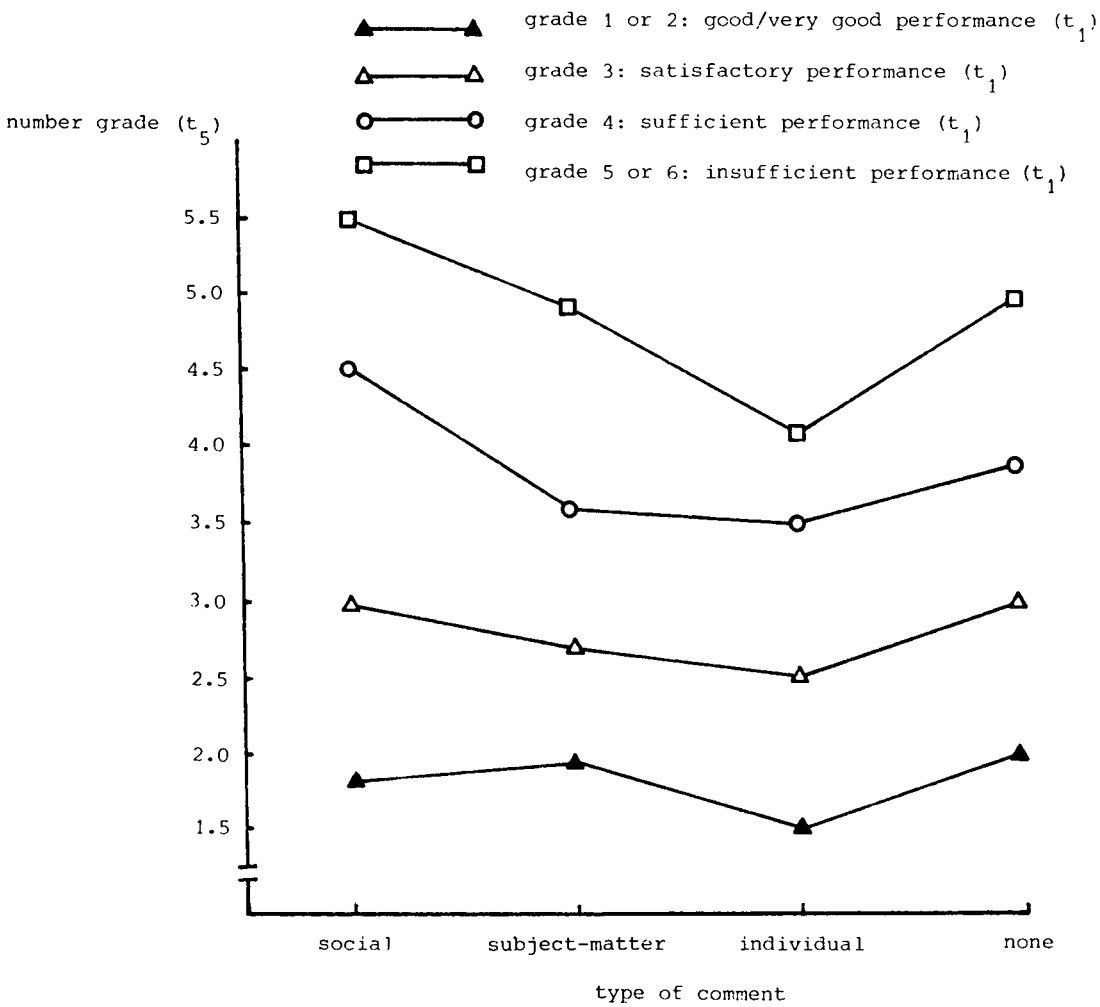

Figure 2. Type of Comment × Prior Performance Level interaction for semester grade in mathematics at t_5.

level and experimental condition. As on the individual level (see Table 5), there were significant main effects for performance level and for type of comment, as well as a significant interaction effect, confirming our hypotheses (see Table 8). Estimates for effect size were also similar for type of comment and for the interaction, whereas those for prior performance level increased considerably. Similar phenomena are frequently found when changing the unit of analysis (Helmke, 1983), using another kind of information, and they should not be confused here. In the present case, this result means that earlier level of performance can predict current performance on the group (class) level with very high probability, higher than on the individual level.

Discussion

The results demonstrate that written teacher comments about grades on examinations are not a simple method for optimizing learning and achievement (Stewart & White, 1976). In past research, only the global efficacy of commenting on exam grades has been looked at, and the interaction between a student's level of performance and the type of teacher comment has not been analyzed. The present study shows that main effects are detectable but that they have to be interpreted differentially in the light of the interactions between comment type and performance level. With respect to the aptitude–treatment interaction, the effects of comments were found to depend on prior level of performance for almost all dependent variables.

The following differential effects of teacher comments were observed: (a) Socially oriented comments clearly affected low-performing students negatively and medium- and high-performing students somewhat positively or not at all. This type of comment further accentuated between-student differences for a variety of motivation, performance, and personality variables. (b) Subject-matter-oriented comments tended to yield generally small positive effects for all students, with no greater advantage for a specific level of performance.

Table 6
Analyses of Variance (ANOVAs) on Posttest Attitude and Personality Measures Using Treatment and Prior Performance Level as Variables

Dependent variable and posttest	ANOVA effect						
	Treatment		Performance level		Interaction		
	F	ω^2	F	ω^2	F	ω^2	MS_e
Internality							
t_5	3.82*	0.041	7.18*	0.092	2.91*	0.024	32.11
t_9	2.71*	0.022	5.79*	0.054	1.44		39.89
Powerful others externality							
t_5	4.66*	0.051	7.15*	0.073	3.17*	0.034	39.36
t_9	2.90*	0.032	3.99*	0.031	1.02		43.13
Chance control							
t_5	2.71*	0.031	1.82		0.98		34.51
t_9	2.45		2.58		0.88		37.17
Attitude toward school							
t_5	4.05*	0.063	2.54*	0.031	2.99*	0.042	6.94
t_9	2.93*	0.031	1.07		0.81		9.77
Test anxiety							
t_5	2.75*	0.023	6.03*	0.084	2.37*	0.032	18.81
t_9	2.89*	0.024	3.71*	0.041	1.12		28.95
Manifest anxiety							
t_5	3.61*	0.031	1.72		1.13		17.00
t_9	1.32		1.53		0.76		25.85

Note. The ω^2 is given only if statistical significance was reached.
* $p < .05$.

(c) Likewise, individually oriented comments showed positive effects, with low performers profiting the most.

Such differential effects of comments on grades can be obtained only for variables that can change within the period of time under investigation. After teacher comments stopped, the effects on grades and cognitive-motivational variables soon disappeared. Only scope-specific personality and attitude variables (aspects of locus of control, attitude toward school, and test anxiety) showed lasting effects. There was a long-term increase in negative attitude toward school and powerful others locus of control in Group 1 (socially oriented comments) and a long-term decrease in chance control and test anxiety as well as an increase in internality in Group 3 (individually oriented comments).

Sizes of almost all statistically significant effects were, according to Cohen's (1977) criteria, medium or small (up to 12% of variance). Elawar and Corno (1985) reported higher effects of teacher comments on student homework, which are more frequent and more continuous feedback on achievement. Effect sizes in this study did, however, reach values that should not be neglected in practice; in contradiction to Stewart and White's (1976) study, teacher comments on exams were demonstrated to have effects with educational significance, especially when their interaction with student performance level was considered. As expected, the most distinct effects were found for situation-specific cognitive-motivational variables, especially for the subjective expectancy of improvement.

The "pure" comments applied in the present investigation probably do not correspond to real school situations, in which mixed types of comments are more frequent. The present results suggest, however, that teachers should emphasize different aspects and contents, depending on the student's prior performance level. Commenting on examinations as well as

Table 7
Means and Standard Deviations of Exam Grades (Intermediate Tests) and Analyses of Variance Using Treatment as the Variable

Test	Group 1		Group 2		Group 3		Group 4				
	M	SD	M	SD	M	SD	M	SD	F	ω^2	MS_e
t_2	3.4	1.2	3.5	1.3	3.3	1.1	3.4	1.3	1.03		5.35
t_3	3.3	1.1	3.8	1.1	3.1	1.1	3.3	1.3	7.19*	.05	1.33
t_4	3.4	1.4	3.2	1.3	2.9	1.3	3.4	1.5	4.77*	.03	1.98
t_6	3.5	1.3	3.3	1.1	3.0	1.0	3.5	1.5	3.41*	.02	1.87
t_7	3.3	0.9	3.5	1.2	3.4	1.3	3.4	1.2	1.42		4.38
t_8	3.4	1.2	3.4	1.4	3.3	1.2	3.4	1.3	0.87		5.13

Note. The ω^2 is given only if statistical significance was reached.
* $p < .05$.

Table 8
Analysis of Variance (ANOVA) on First Posttest Reported Grade Using Treatment and Prior Performance Level as Variables

Source	*df*	*F*	ω^2
Treatment (T)	3	66.36*	0.056
Performance level (P)	3	1,047.72*	0.905
T × P	9	10.00*	0.023
Residual (MS_e)	36	0.02	

Note. Class means of reported grades were used as the dependent variable for the ANOVA.
* $p < .05$.

on student homework (Elawar & Corno, 1985) may be most effective. In contrast to Elawar and Corno's, our results show that written comments should be content specific and take into account a student's performance level and perhaps should also take into account a student's concept of his or her own competence. Otherwise, our findings show that teacher comments do not produce only positive effects.

References

Atkinson, J. W., & Birch, D. (1970). *The dynamics of action.* New York: Wiley.

Batlis, N. C. (1978). Relationships between locus of control and instrumentality theory predictor of academic achievement. *Psychological Reports, 43,* 239–245.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation. Design and analysis issues for field settings.* Chicago: Rand McNally.

Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework. *Journal of Educational Psychology, 77,* 162–173.

Fittkau, B., & Langer, H. (1974). Auswirkungen schriftlicher Ermutigungen unter Klassenarbeiten auf Angst und Leistungen der Schüler [Effects of written encouragement on exams on students' anxiety and performance]. *Psychologie in Erziehung und Unterricht, 21,* 15–21.

Grobe, R., & Hofer, M. (1983). Kognitiv-motivationale Korrelate von Schulnoten [Cognitive-motivational correlates of school grades]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 15,* 292–316.

Heckhausen, H. (1977). Achievement motivation and its constructs: A cognitive model. *Motivation and Emotion, 1,* 283–329.

Helmke, A. (1983). *Schulische Leistungsangst* [Test anxiety in schools]. Frankfurt, FRG: Lang.

Henson, S. (1976). Expectancy, beliefs, ability, and personality in predicting academic performance. *Journal of Educational Research, 70,* 41–44.

Kahl, T. N. (1977). *Unterrichtsforschung* [Research of teaching]. Kronberg, FRG: Scriptor.

Klauer, K. J. (1982). Bezugsnormen zur Leistungsbewertung [Reference norms for achievement evaluations]. In F. Rheinberg (Ed.), *Bezugsnormen zur Schulleistungsbewertung* (pp. 71–89) [Reference norms for achievement evaluations in schools]. Düsseldorf, FRG: Schwann.

Krampen, G. (1979). Zur prognostischen Bedeutung kognitiv-motivationaler Effekte von Zensuren in einer Deutscharbeit bei Hauptschülern [Predictive significance of cognitive-motivational effects of grades on a German exam among students]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 11,* 250–260.

Krampen, G. (1984). Handlungstheoretische Perspektiven zur Analyse politischer Partizipation und Apathie [Action-theoretical perspectives on the analysis of political participation and apathy]. In H. Moser & S. Preiser (Eds.), *Umweltprobleme und arbeitslosigkeit* (pp. 27–42) [Environmental problems and unemployment]. Weinheim, FRG: Beltz.

Krampen, G., & Lehmann, P. (1981). Zum Vorhersagewert der kognitiven Motivationstheorie für schulische Zensuren, Verbesserungsmotivationen und Leistungsveränderungen [Predictive significance of cognitive motivation theory for school grades, improvement motivation, and improvement]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 13,* 194–206.

Mathis, B. C., Cotton, J. W., & Sechrest, L. (1970). *Psychological foundations of education.* New York: Academic Press.

Misanchuk, E. R. (1977). A model-based prediction of scholastic achievement. *Journal of Educational Research, 71,* 30–35.

Mitchell, T. R., & Pollard, W. E. (1973). Instrumentality theory predictions of academic behavior. *Journal of Social Psychology, 89,* 34–45.

Page, E. B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *Journal of Educational Psychology, 49,* 173–181.

Pekrun, R. (1983). *Schulische Persönlichkeitsentwicklung* [Character development in school]. Frankfurt, FRG: Lang.

Rheinberg, F. (1980). *Leistungsbewertung und Lernmotivation* [Performance evaluation and motivation for learning]. Göttingen, FRG: Hogrefe.

Rheinberg, F., Kühmel, B., & Duscha, R. (1979). Experimentell variierte Schulleistungsbewertung und ihre motivationalen Folgen [Experimental variation of performance evaluation and its motivational consequences]. *Zeitschrift für Empirische Pädagogik, 3,* 1–12.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*(1, Whole No. 609).

Schmidt, W. (1978). Darstellung und Verwendung der Leistungsmessung und Leistungsbeurteilung in der Schule [Description and use of performance measurement and performance evaluation in schools]. In E. Stephan & W. Schmidt (Eds.), *Messen und Beurteilen von Schülerleistungen* (pp. 153–171) [Measuring and evaluation students' performance]. Munich, FRG: Urban & Schwarzenberg.

Schwarzer, R. (1979). Sequentielle Prädiktion des Schulerfolgs [Sequential prediction of success in school]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 11,* 170–180.

Stewart, L. G., & White, M. A. (1976). Teacher comment, letter grades, and student performance: What do we really know? *Journal of Educational Psychology, 68,* 488–500.

Wieczerkowski, W., Nickel, H., Janowski, A., Fittkau, B., & Rauer, W. (1975). *Angstfragebogen für Schüler (AFS)* [Anxiety Questionnaire for Students]. Braunschweig, FRG: Westermann.

Received June 5, 1985
Revision received November 10, 1986
Accepted December 1, 1986 ■