DOCUMENT RESUME

ED 125 907 SE 021 153

AUTHOR Good, Thomas L.; Grouws, Douglas A.
TITLE Process-Product Relationships in Fourth Grade Mathematics Classrooms.
INSTITUTION National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Oct 75
GRANT NEG-00-3-0123
NOTE 123p.; Figures at end of document marginally legible

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS Achievement; *Classroom Environment; *Effective Teaching; Elementary Education; *Elementary School Mathematics; *Instruction; Mathematics Education; *Research; Student Participation; Student Teacher Relationship

ABSTRACT

This volume is the final report of a study of teacher effectiveness. the research, sponsored by the National Institute of Education, was designed to relate student achievement in fourth-grade mathematics to teachers' use of various strategies and to classroom environmental factors. Six clusters of variables were found to be strongly associated with teacher effectiveness: (1) student initiated behavior, (2) whole-class instruction, (3) general clarity of instruction, (4) a nonevaluative and generally relaxed environment, (5) higher achievement expectations, and (6) classrooms that were relatively free of major behavior disorders. This report includes the following analyses of data obtained from the Iowa Test of Basic Skills and from classroom observation: (1) correlational and analysis of variance results linking process measures with mean residual achievement; (2) discriminant analyses of significant process relationships (testing the extent to which process measures discriminant high and low teachers); (3) data describing where teachers obtain their gains; and (4) correlation of process data with classroom climate scores and data describing the relationship between residual mean achievement scores and classroom climate scores. (SD)

Process-Product Relationship in Fourth Grade Mathematics Classrooms

Thomas L. Good
Douglas A. Grouws
University of Missouri-Columbia

October 1975

Final Report of
National Institute of Education
Grant NEG-00-3-0123

Process-Product Relationships in Fourth Grade Mathematics Classrooms*

Thomas L. Good, Principal Investigator
Douglas A. Grouws, Co-Principal Investigator

*This is the final report of the National Institute of Education grant NEG-00-3-0123. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement from that office should be inferred.

The authors gratefully acknowledge the work of Terrill and Sally Beckerman who collected the observational data and the assistance of Dr. Ernest Hilderbrand, David Hentchel, and Jim Worstell in analyzing the data. We are also in debt to Sherry Kilgore for her capable assistance in the typing and general preparation of the final report. Secretarial support and general assistance was provided by the Center for Research in Social Behavior, University of Missouri, Bruce J. Biddle, Director.

Abstract

This is the final report of N.I.E. project NE-G-00-3-0123 (originally titled Multiple Criteria of Teaching Effectiveness). Fourth grade teachers were identified for observation on the basis of students' residual mean performance on the mathematics section of the Iowa Test of Basic Skills. Observation and data analysis focused upon teachers who were stable and relatively high or low in terms of their students' performance across consecutive years.

Included in this report are:

(1) Correlational and analysis of variance results linking process measures with mean residual achievement.

(2) Discriminant analyses of significant process relationships (testing the extent to which process measures discriminate high and low teachers).

(3) Data describing where teachers obtain their gains (i.e. where do relatively effective teachers obtain their gains--from high, middle, or low achievement students?).

(4) Correlation of process data with classroom climate scores and data describing the relationship between residual mean achievement scores and classroom climate scores.

Historically, teacher effectiveness research has been a popular but unproductive activity. Most teacher effectiveness research has employed criteria of teaching effectiveness that lack validity or generality. The low productivity of teaching effectiveness research has frequently been documented (Morsh and Wilder, 1954; Stephans, 1967; Rosenshine and Furst, 1971; Dunkin and Biddle, 1974). Mitzell and Gross (1958) reviewed several studies that had used pupil growth criteria as an index of teacher effectiveness and noted that authors of such studies provided no assurance that effective teachers are "effective" in any absolute sense. At best teachers' relative effectiveness was argued on the grounds of effectiveness in comparison to other teachers in the sample (and often a very small sample).

However, even the criterion of relative effectiveness was often weakened by the tendency to select first year teachers for such studies. Thus, in many cases effective teaching studies involved an examination of first year teachers who were but superficially superior to other first year teachers. Understandably, the collective results of such studies are conflicting and weak information.

How to Pick a Sample

Research on teacher effectiveness in which effectiveness is gauged through child gain measures usually depends on multiple regression or covariance techniques to adjust class gain scores for varying levels of initial child performance. At first glance, this appears to be a logical method of looking at the effects of teachers on student performance. Regression techniques statistically control differences in initial (prior to instruction) student behavior; thus, it is argued that the differential student gains in classroom A "must be due to teacher A" because child differences are statistically removed. However, teacher A in September may have 20 of the 30 third graders above grade level achievement while in classroom B the teacher may have only 6 of 30 students above grade level. Statistical controls are not related, in any direct way, to important differences between such classrooms (the incidental learning and motivational climate associated with the grouping of good students in the same classroom, for instance). Regression analysis is the appropriate way to approach this problem; however, efforts should initially be made to adjust the initial differences between teachers as much as possible through naturalistic means. Classroom conditions that vary greatly cannot be equated through statistical means.

Unless teachers are teaching similar students with similar backgrounds (SES of pupils in the classrooms under comparison is homogeneous), it is not possible to argue that certain teachers (on the dependent measure in question) are more effective than others. To study relationships between classroom behavior and student achievement it is necessary to compare teachers who teach in at least quasi-equivalent circumstances. This can be done by deliberately matching classrooms for relevant student characteristics or by capitalizing on such matching when it occurs naturalistically in schools practicing random grouping. However, it is possible to identify but few studies which have employed such methodology to assure that the teachers under investigation were teaching under quasi-similar conditions (Torrance and Parent, 1966; and Soar, 1966).

During the planning of this project, the results from Jere Brophy and Carolyn Evertson's program of research were just beginning to emerge. They had improved upon the general methodology of teacher effectiveness research by reducing systematic differences between classes on variables not included as co-variables by using different regression formulas for sex, SES, and pretest level (Brophy, 1972). Significantly, they used all teachers in the school district who taught at target grade levels and who met selection criteria. Hence, they constructed a sample of teachers with known (and varying) effectiveness prior to the collection of observation data.

In general, the Texas Teaching Effectiveness project had demonstrated that teachers do differ in their ability to influence student achievement scores and that some teachers were stable in their effects on students over time. Hence, we were interested in replicating and extending this line of research with a different population.

We were especially interested in conducting the study because we had access to student achievement records from a large school district that "skirted" a large metropolitan area in the midwest. Comparatively, there was but little SES variance across schools. The district served a largely middle class population. Thus, through naturalistic and statistical controls it was possible to judge the relative effectiveness of a large number of teachers who were teaching similar students with similar materials (the same math series was used in all schools). The relative homogeneity of the population weakens the external validity (the application of these results to other settings) of the study, but, correspondingly, this condition provides a more rigorous test of teacher effects on students' performance (as measured by standardized achievement tests).

Method for Picking Observation Sample

The data unit for this study consisted of individual students' scores on the Iowa Test of Basic Skills. The test scores from the fall of the third grade were used as pre-scores for the third grade and tests given in the fall of the fourth grade were used as pos scores. Similarly, the tests administered in the fall of the fourth grade were used as pre-scores for the fourth grade and post scores were obtained by fall testing in the fifth grade. These data were compiled for fall testing in 1972, 1973, and 1974.

Grade level equivalents rather than raw scores were used since these are more normalized measures likely to contain less error variance than the raw scores. Data were used from the following subtests: vocabulary, reading comprehension, total language skills, total work study, mathematics concepts, mathematics problem solving, total mathematics, and a composite score.

Residual gain scores were computed for students on each subtest by using the student's score on the pre-score subtest as a covariate (using a linear model where $g = y - (a + bx)$. Residual gain scores were computed within year and within grade level (third and fourth). Data for teachers were then compiled by computing a mean residual gain score (from the scores for students in each of their two classes) for year 1 and year 2.

The initial sample included almost all third and fourth grade teachers in the school district. The testing procedure employed in the school district called for early fall testing. Unfortunately, the school district tested at every other grade level so it was impossible to construct residual scores from extant data without additional testing.

Fall testing offers advantages as well as disadvantages. It can be argued that fall testing improves the objectivity of teacher administered standardized achievement tests as teachers are testing previous instruction not their own. However, procedurally this arrangement constitutes a major data preparation problem. As we mentioned above, to compute a third grade teacher's residual mean score it is necessary to find a given student's achievement score in that teacher's room in the fall, and also to find that student's score the following fall (that student of course may be in one of four or five fourth grade rooms).

To further complicate matters teachers may share students (with another teacher) for instruction in math or reading. Hence, even though a student's name appears on Ms. X's Iowa achievement roster he may or may not be instructed by her during language arts or mathematics activities. Approximately one-half of the teachers in this sample shared students for instructional purposes (mathematics, 45%; reading, 55%). Sharing in some cases involved the transfer of but 2 or 3 students, but, in some cases, it involved over a third of the students.

Hence, the fact that we had to test children to obtain data essential for computing residual scores prevented us from conveniently selecting the sample on the basis of available data. However, it did alert us to the fact that many teachers were sharing students and the need for obtaining class rosters on the basis of those students taught by a teacher in mathematics and reading.

The task proved to be an administratively complex one. However, teachers who shared students were identified and classroom rosters were adjusted accordingly. Despite the careful monitoring we imposed on the data, we regard the stability data with a degree of skepticism. There is a great deal of room for slippage. For example, some teachers who share students for a couple of weeks may not conceive of this as sharing and hence not report it.

In general, we regard all reports of teacher stability (on student achievement) with some suspicion if they are not supported with rosters obtained by observers who can substantiate the fact that students are receiving their mathematics instruction from a particular teacher. This has become especially important during the past two or three years as more and more school organization plans call for students to have contact with more than one teacner.

With the limitations mentioned above in mind, the data provide support for the contention that some teachers are stable in their effects upon students. The median year to year correlation across residual gain scores on all subtests of the Iowa Achievement test of Basic Skills was approximately .20.

Originally we had planned to draw a sample of teachers who were relatively high and stable on at least a couple of cognitive measures (e.g. composite, total math) and an affective measure. But, as we shall discuss below, stability problems plagued us here as well.

The Rabinowitz-Rosenbaum Teacher Rapport Scale was selected for use in this study because of its conciseness, ease of administration, and face validity. However, in addition to its efficient administration properties the test has several commendable features. The test for example was high reliability (.89 estimated by analysis of variance procedures) and the test can be administered to children in a group setting (Rabinowitz and Rosenbaum, 1958). Also, the test has been found to correlate with measures of teacher behavior (such as coded, manifest hostility) and to produce a global halo estimate of student attitude toward teacher and class (Kleinfeld, 1972; Rabinowitz and Rosenbaum, 1958). A copy of this instrument appears in Appendix B.

However, teacher stability on the classroom climate scale proved to be no better (across consecutive years) than did the stability on the Iowa Achievement Test. The correlation of rankings from two separate groups of students (spring 1973 and spring 1974) was .22. Hence, different student groups saw teachers and classroom environments somewhat differently.

But as we have pointed out elsewhere (Good and Grouws, 1975) the scores in both years were generally high and the variance quite restricted. Still, teachers at the extreme ends of the distribution tended to maintain their relative rank. Classrooms and teachers rated especially high one year tended to be rated as relatively high the following year. Similarly, teachers who were rated very low one year tended to be low the following year. The pattern for low "climate" teachers was similar to the pattern observed for teachers who had low mean residual scores on the Iowa Test of Basic Skills. Teachers toward the bottom of the distribution one year stayed in the bottom fourth of the distribution the following year.

Given the fact that teachers' total residual score did not correlate at all with classroom climate and that no subtest residual correlated higher than .20 with climate, it was impossible to use multiple criteria for selecting an observation sample. For the group of teachers as a whole, there essentially was no relationship between classroom climate scores for a teacher and teachers' mean residual gain (on the Iowa Composite score: total score). Interestingly, there was a strong correlation (.50) between teachers' residual mean gain score on total math and classroom climate during the year of observation. Possible explanations to account for these differences will be presented later.

In review, three findings from the climate (affect) scale have been reported. First, in the large sample (using 104 classrooms) there was no relationship between composite residual mean score (residual year 1) for teachers and classroom climate scores. Second, over consecutive years there was but little teacher stability on climate score (.22). Third, in the intensive observational study (41 classes) there was a correlation between climate and residual gain.

Additional problems in selecting an observational sample were caused by the fact that a few teachers had moved out of the district and a few teachers that had been self-contained teachers were now team teaching. Furthermore, teachers' participation in the observational phase of the study was on a volunteer basis (roughly 75 per cent of the teachers contacted for participation via letter agreed to do so). A small honorarium of $50 was paid to each teacher.

To maximize the utility of the data we decided to focus the observation upon teachers' performance in mathematics. This decision was reached because (1) teachers available for observation demonstrated slightly more stability in mathematics than reading and (2) we felt more prepared to do a detailed analysis of mathematics than reading (the co-principal investigator is a specialist in mathematics education); (3) subsequent data from the Brophy-Evertson paradigm suggested that teaching effects were contextual (different patterns of effectiveness had been noted in low and high SES teaching situations). Hence, we felt an intensive examination in a single subject area would be a more powerful probe, and for the reasons given above we chose to focus on mathematics.

This strategy made it possible to identify nine teachers who were relatively effective and stable on total math residual scores across two consecutive years (that is they were in the top one-third of the sample across two years) and nine teachers who were relatively ineffective and stable across two consecutive years. Furthermore, these teachers all basically taught mathematics as a whole class activity. Other teachers included in the observation sample taught mathematics with students assigned to groups, and a few teachers taught mathematics in an individualized mode.

Thus it is possible to compare the effects of teaching the whole class vs. teaching groups of students (too few individualized classrooms were observed to compare with other modes) as well as process differences between teachers with known (and varying) records of effectiveness. However it should be realized that teachers who taught the whole class might have 1 or 2 students working independently and one or two teachers who taught mathematics in a group format might have one or two students working independently.

Observational Method

Observational data were collected by two trained observers (both certified teachers) who worked full time on the project and who lived in the target city. (The data were being collected in a city some distance from our own location.) Prior to the collection of observation data the observers went through a three and one-half week training program that involved the coding of written transcripts, videotapes, and live observations in fourth grade mathematics classrooms.

Video codings and transcript codings allowed for both reliability and validity checks (i.e. Did the coders agree and were their judgments correct?). In general, the reliability data are quite good especially for the Brophy-Good Dyadic System, the basic coding instrument in the study. All coding distinctions made with this instrument had reliabilities above .80 (usually .90 or higher) with the exception of praise and criticism (.70 and .75). Most of the coding distinctions in the dyadic are straightforward (e.g. either the teacher approaches the child or the child approaches the teacher) and involve but little interpretation.

Coder agreement on the high inference scales was acceptable although, generally, not as high as on the low inference Dyadic. Reliability data and scale definitions appear in Appendix A.

Four basic sets of information were collected in the study. First, time measures were taken to describe how mathematics instructional time was utilized. In addition to descriptive purposes, these categories were designed to facilitate the testing of several hypotheses that flow from experimental research on time variables in mathematics instruction (e.g. the ratio of time spent in development vs. practice activity).

The second set of codings were low inference descriptions of teacher-student interaction patterns (the Brophy-Good Dyadic System described above). The third set of data were high inference variables drawn from the work of Emmer (1973) and Kounin (1970). Brief definitions of all variables involved in the study can be found in Appendix A along with copies of all coding sheets. Extended definitions of terms and coding examples can be found in the original sources (Brophy and Good, 1970; Kounin, 1970; and Emmer, 1973).

The fourth type of data coded were checklists that were used to describe materials and homework assignments. These checklists can also be found in Appendix A.

In summary, coding scales were designed to describe: the distribution of time in mathematics instruction; the ways in which teachers and children interacted during mathematics instruction (with both low and high inference descriptions); and the general use of material and resources (including out of class assignments in the form of homework).

During October, November, and early December observational data were collected in the rooms of the 41 participating teachers during mathematics instruction. Each observer made (roughly) one-half of the visits for each teacher. Hence teachers were seen at least a couple of times by each observer. Each teacher was observed from 5 to 7 times for an entire mathematics period. Observers arrived a few minutes prior to the time that mathematics instruction was to begin (partly in order to code the amount of transition time necessary to switch from one activity to another) and stayed until mathematics instruction terminated.

Process (observational) data collected during this time period were subsequently analyzed with classroom mean residual scores on total mathematics (on the Iowa Test of Basic Skills) collected during that _same_ year. Product measures (Iowa Test of Basic Skills) were administered in October 1974 and April 1975, and classroom residuals were computed from these two data sets. However, before presenting process-product relationships, it is important to briefly discuss a few interpretation difficulties that are imposed by using a variety of process measures.

A Note on Interpreting Results

In reacting to the data that follow, it is important to realize that different metrics are being utilized. First, time measures are reported in terms of a mean score across all observations. The means for transition, the second variable listed on Table 1, are correctly interpreted as follows: the 9 high teachers, as a group, averaged two minutes and forty-five seconds in transition each observation; whereas, the 9 low teachers, as a group, averaged three minutes and fourteen seconds per observation.

Second, low inference behavioral measures are reported in terms of rates per hour. To control for unequal length and number of observations, the frequency of each behavioral measure was divided by the number of minutes in the observation and multiplied by sixty. The first variable reported from the Brophy-Good Dyadic is Student Asks Question (see Table 1, top of page 2). The means here are interpreted as follows: students in the 9 high classrooms initiated, on the average, 2.84 questions per hour; whereas, students in the low classrooms initiated, on the average, 2.07 questions per hour.

High inference measures are represented by a mean score across observations. Five high inference variables (organization, alerting, accountability, classroom climate) were rated once at the end of each observation. Hence the rating is a single global estimate, and the means that appear in Table 1 for these variables is the average of the ratings that were made at the end of each observation.

However, several high inference variables (general thrust of homework; student attention; clarity; enthusiasm; average accountability; average alerting; percent of students probably involved; and percent of students definitely not involved) could be measured several times during an observation. Here a mean is computed for separate ratings within an observation and then the mean score for an observation is combined with the mean score of other observations to yield a general mean.

A problem to be aware of when interpreting high inference measures is that the variables are coded on 3, 4, and 5 point scales. Furthermore some of the scales have been reversed so that a higher score does not always mean more of a variable. To interpret the high inference variables, appropriately, the reader is advised to examine scale definitions that appear in Appendix A.

To enhance the interpretability of the data, high inference measures in Table 1 are marked with an asterisk. However, the data in all tables are reported in the following sequence. First, the time measures appear. Then come the high inference measures (the five single rating variables appear first, followed by the variables that might be rated more than once during an observation), and finally the dyadic behavioral codes are reported.

As can be seen, on page 1, Table 1, the first variable listed is the average number of students present during mathematics instruction. This variable is followed by the nine time variables, and the 14 high inference ratings. The remainder of the variables reported in Table 1 are drawn from the Brophy-Good (low-inference) Dyadic Observation System. To reiterate low-inference data are reported in per-hour frequency rates.

Results

Table 1 presents means and significance levels from an analysis of variance computed across process measures for the top 9 and bottom 9 teachers. Table 2 is a condensation of Table 1 and presents only significant or near significant findings from these same analyses. As noted earlier, we view these data to be the most basic and most important, because they describe process measures collected in classrooms of teachers with known levels of effectiveness and demonstrated patterns of stability.

Tables 3 and 4 present the correlations of all process measures with residual achievement scores. Table 3 reports all variables and Table 4 summarizes significant and near significant process-product relationships. Unlike the analysis of variance results in Tables 1 and 2, the correlation data are based upon all the teachers in the observation sample: not just the top nine and bottom nine teachers.

Top Nine and Bottom Nine Teachers

As can be seen in Table 1, those teachers with high residual gain scores (highs) have more students, spend more time teaching the whole class, and spend more time in developmental activities than do teachers with lower residual gain scores (lows). Low teachers in contrast spend more time in transition, spend more time going over homework, and more time in drill activities. The biggest time difference is that highs spend more time teaching the class as an intact group. Although lows basically teach the whole class, they spend more time teaching groups of students or individuals than do highs.

At the end of each observation, coders made five high inference ratings to describe the entire observation. In general, these variables (organization, alerting, accountability, climate, and managerial) do not differentiate the process in high and low classrooms. However, the classroom climate rating does appear to be different in the two groups of teachers. In general, high classrooms (as rated on the climate scale) are seen as more relaxed than are low classrooms.

In contrast, many of the high inference measures that are coded several times during an observation, do show interesting results across the two groups of teachers. For example, high teachers, when they do conduct discussions of homework, tend to place more emphasis upon process than product. Although the difference in means on this particular measure is small, this difference (more process feedback from high teachers) is repeated throughout the data.

Also, it can be seen that students are rated as paying more attention in high classrooms and this occurs despite the fact that highs do not average more accountability and alerting statements than do low teachers. Initially we were surprised that lows engaged in more alerting and accountability behavior than highs, but subsequently we interpreted this finding to be due to the operational definitions of these two variables. More on this later.

Importantly, high teachers present with greater clarity than do lows. This communication facility may be part of the reason why they are successful in teaching the class as a whole.

The low inference data begin with the variable, student asks question. The difference in means on this variable is quite low; however, there is a small trend for students to ask more questions in high rooms than in low rooms. This trend (student initiation) is repeated throughout the data. For example, consider the variable, student calls out answer. Here, again, we see that students in high classrooms are much more likely to call out answers than students in low rooms.

Essentially, the data on page 2 of Table 1 indicate that low teachers ask more questions than do highs. However, a couple of interesting contrasts appear here. As can be seen, lows ask many more process questions than do high teachers. But later we shall see that highs provide students with considerably more process feedback than do lows.

Interestingly, twice as many incorrect responses occur in low classrooms. It can be argued that the large number of incorrect responses is understandable given the higher rate of questioning in low rooms. Still, the number of incorrect responses per se may be so high as to interfere with classroom learning (e.g. large number of incorrect answers to model, much instructional time spent in negating answers, etc.).

Another variable of interest is the tendency for highs to praise correct answers by students more than lows. This is especially interesting because, as can be seen on page 7, lows followed students' answers with praise seven times more frequently than did highs. Apparently, low teachers praised student effort much more than did highs and high praise rates do not relate to achievement (as we shall see in Table 3). Low teachers praised students considerably more frequently than did high teachers. High teachers appear to praise more contingently.

One of the most significant findings in the entire data set appear at the bottom of page 3 and the top of page 4. These data consistently and strongly indicate that students in high classrooms initiate much more contact with the teacher than do students in low classrooms. This is especially the case in work-related interactions. In contrast, low teachers initiate considerably more contact with students than do highs. The consistency of high teacher initiation rates in low classrooms is fully demonstrated on page 4 of Table 1.

Students in high classrooms initiate twice as many work-related contacts with teachers as do students in low classrooms. Thus, despite the fact that the total number of dyadic contacts in high and low classrooms is quite similar, the pattern across groups is strikingly different.

Proportionately, most of the private dyadic contacts between student and teacher that occur in high classrooms are due to student initiation. In contrast most of the private interaction that occurs in low classrooms are due to teacher afforded interaction.

Interestingly, despite the high general rates of teacher praise in low classrooms, lows praise students less in student initiated work contacts. Hence, as can be seen at the bottom of page 3, Table 1, lows do not encourage students when students do approach them in work-related situations. These data would suggest (at least indirectly) that students do not approach teachers in low rooms because teachers do not want them to do so.

Two additional findings of interest appear on page 5 of Table 1. First, lows proportionately ask more direct questions than do highs and second, lows proportionately ask more process questions (twice as many) than do high teachers. However, as we see on page 6 of Table 1, lows provide proportionately less process feedback than do high teachers. In general highs provide twice as much process feedback as do lows.

Table 2 presents, in abbreviated form, the highlights of Table 1. In review, the data suggest that high teachers spend more time teaching the whole class as a group and create a more relaxed, more interesting classroom climate than do lows. However, at the same time they present with more clarity and appear to strike an interesting combination of engaging the entire class on task while maintaining a relaxed environment. High teachers' ability to present material clearly and to keep students attention is no doubt part of the reason that teaching the whole class (as a unit) works for them.

In a relative sense, it appears that low teachers are poor classroom managers. This can be seen in the fact that they raise more discipline questions, as well as warning and criticizing students more frequently than do highs. No doubt such behavioral criticism and warnings work against the creation of a relaxed learning environment. Higher alerting and accountability scores in the context of frequent criticism and warnings may take on an adversive quality (e.g. students may fear failure).

In general, we see that students initiate much more interaction in high classrooms. This finding is coupled with an interesting form of complimentary teacher behavior: the provision of process feedback. Highs do not ask many process questions. Apparently, these teachers prefer to provide process feedback when students are ready to listen to process explanations.

In public settings, the teacher spends proportionately more time on developmental activities (explaining the process publicly to all students) than do lows, but keeps the ball moving rather than frequently calling upon students. When highs do call on students in public settings, they are most likely to ask product questions . . . presumably followed by a quick student response.

The public presentations of highs are interrupted by student questions and call-out responses more frequently than in low classrooms. Hence, these may be mechanisms through which teachers allow students to get feedback in public settings.

Process-Product Correlations

Table 3 presents the same variables as discussed in Table 1. But Table 3 describes correlations between process measures and students' achievement. There is one additional difference. Table 1 is based upon 18 teachers (the top 9 and bottom 9); whereas, Table 3 is based upon all teachers in the observation sample. Thus, it is possible that some of the relationships that appear to be sharp differences between highs and lows in Table 1 may become blurred in Table 3 because teachers in the middle range of effectiveness show a mixed pattern on the variable (some are like the highs; some are like the lows). However, it is also possible that using the full range of teachers might strengthen the importance of some variables. The sample size* upon which correlations are based is 33.

We now turn to discussion of the data in Table 3. When the sample as a whole is considered, the time that the teacher teaches the whole class appears to be considerably less important. This is largely because most teachers falling in the middle range effectiveness in this study do not teach the class as a whole (see Table 11). Interestingly, time spent in review relates positively to achievement and is a more interesting variable than it appeared to be before.

Classroom climate shows the same pattern as observed in Table 1. A relaxed climate correlates moderately with student gains. Also, it can be seen on page 1 of Table 5 that number of students present for instruction still relates positively to achievement gains.

A new variable of interest in this data set is the high inference rating: teacher presentation. Essentially this scale describes the percent of time that the teacher is presenting material. The data suggest that high rates of teacher presentation of information is negatively correlated with student achievement.

---

*Residual gain scores were not computed in some of the individualized rooms because of high numbers of third or fifth grade students.

Accountability and alerting ratings appear to be marginally related to student gain scores. Measures of student attention and involvement show no relation to student achievement.

The data describing frequency and type of questions that teachers ask students generally indicate that more frequent questioning correlates negatively with student gain scores and this is especially so when students give wrong answers or when they fail to make a response. We suspect that curvilinear data analyses will reveal that frequency of questioning, in math classrooms, pays off up to a point but after that point additional teacher questioning wastes time.

Interestingly, students' answers that are partially right show a very moderate correlation with achievement. This finding in and of itself means very little. However, on page 3 of Table 3, we can see that part right responses followed by sustaining feedback are correlated with student achievement. This may suggest that partially correct answers indicate a teacher's ability to phrase questions that appropriately bridge the gap between what a student does know and new concepts that the student is just beginning to form.

Although the strength and direction of teacher initiated and student initiated correlations vary from variable to variable, it is clear that teacher initiated contacts generally are negatively correlated with student gain; whereas, student initiated work contacts are related to student achievement gains.

Three findings on page 6 of Table 3 appear minimally interesting. First, a higher percentage of correct student responses to teachers' public questions relates positively to student achievement. Second, students' attempts to answer questions are weakly but positively related to student gain.

This can be seen in the variable, wrong responses/wrong responses plus no response. Third, these data (as in Table 1) support the value of process feedback. However, process feedback appears to be useful in public settings and harmful in private settings. But recall that this data is based upon using all teachers in the sample. Perhaps some of the teachers do not provide process feedback appropriately (e.g. they use too much time doing so) and hence private process feedback does not enhance achievement. This interpretation is similar to data presented by Brophy and Evertson (1974) who report that high SES children benefit from brief feedback.

Table 4 summarizes in profile form the highlights of Table 3. In review,we see that teachers' mean residual gain scores correlate with teachers' affective mean score (the climate inventory). Also, highs have more students in their classroom and spend more time in review. More effective teaching appears to take place in a relaxed classroom environment and in a climate in which students feel free to call out answers to academic questions. Students in high rooms answer a higher percentage of questions correctly and, further, high students seek out the teacher. In low classrooms, teachers seek out students. Interestingly, the frequent use of discipline questions correlates negatively with student gain.

The data in Table 3 generally provide the same pattern as the data reported in Table 1. Two new variables of possible importance (review and presentation) are identified and the importance of one variable is qualified (time teaching the whole class).

Top Three and Bottom Three Teachers

Table 5 presents yet another way of looking at the question we have been asking: what process behavior correlates with student learning performance? Table 5 presents the results of an analysis of variance run across the top three and bottom three teachers. Hence, these results report comparisons across the highest and lowest classes in the sample. In general, these data do not contradict or extend the data that we have reviewed above.

Again, we see that highs teach the class as a class more frequently than do lows and that they are successful in constructing a more relaxed classroom atmosphere. Lows engage in more alerting and accountability behaviors than do highs. Highs ask fewer discipline questions and fewer process questions than do low teachers. However, highs are more likely to provide process feedback to their students than are low teachers.

In general, highs are less likely to stay with students by providing sustaining feedback when students do not make a response or indicate that they do not have to know how to respond. In such situations, high teachers are less likely to repeat the question, to give a clue, or to expand upon student responses. However, as previously noted, the 9 high teachers were more likely to stay with students when they give a partially correct response. The means for the top 3 and bottom 3 teachers are in the right directions (highs sustained 3.40 part right responses per hour; whereas, lows sustained 1.60 per hour). But the p-value does not approach significance and the behavioral frequency of part right responses is low. Still, given that these teachers are proportionately asking more product questions, the strategy appears to make sense. That is, students are asked questions that they either know or do not know and when they indicate that they cannot respond the teacher moves on to someone else or provides a process explanation.

Lows are much more likely to praise students than are high teachers. This is strong evidence that excessively high praise rates are not necessarily associated with student learning. We suspect that in this sample the high praise rates are indicative of an inability to cope with managerial demands of the classroom. That is, teachers through the use of praise are attempting to "buy off" and control students: a strategy that does not appear to work. Part of the problem here may reside in the fact that teacher training programs have been encouraging teachers to increase their use of praise. Some teachers respond to this plea indiscriminantly rather than making praise contingent upon good performance. It is clear that high praise in and of itself is not necessary for learning and may be detrimental (e.g. may communicate low expectations). Data in Table 5 also make it clear that low teachers have more discipline problems. This can be seen in the fact that they ask more discipline questions, give more behavioral warnings, and more behavioral criticisms. The tendency to criticize the behavior of students even spills over into the criticism of academic work. It can be seen in Table 5 that lows are more likely to criticize students when they initiate work contact with students.

Perhaps this is one reason why teacher afforded contact with students is negatively associated with student achievement. That is, if students are unduly concerned about receiving praise and/or criticism from the teacher, the presence of the teacher moving around the room may have a detrimental affect upon student performance.

Once again we see the importance of student initiated work contacts and the relatively detrimental effects of teacher afforded contacts with students. Presumably, in classroom situations that involve children from middle class homes (at least in a focused academic subject like mathematics) it is better to allow students to approach the teacher when they need help. Furthermore, once again process feedback appears to be important in facilitating student progress even though process questions are not.

Process Measures Related to Climate Scores

We have spent considerable time describing the process milieux of classrooms that are characterized by relatively high and low performance in mathematics achievement. However, it is also possible to talk about process measures that differentiate relatively high and low classroom climates (as reported by students). The data in Table 6 report the correlation of process measures (the same process measures reviewed in Table 1 and Table 3) with teachers' affective means (the score on the climate inventory). What process variables are associated with teachers who obtain more favorable attitudinal responses from students?

We see that total time in mathematics, time spent going over homework and review time correlate with affective attitudes of students. Perhaps general review and review of homework provide structure and help students to know precisely what they are to do. This direction or structure may enhance general comfort and security. It may also directly communicate to students that the work they do is worth attention. In general, most time measures correlate with positive climate scores (the exceptions are development and drill but neither correlation is very strong).

At the bottom of page 1 of Table 6 it can be seen that enthusiasm correlates strongly with students' affective responses. Indeed, it relates more strongly to climate scores than does any other process measure. Teachers who are more enthusiastic get higher climate scores from students. Also, it can be seen that teacher alerting statements are associated with high affective responses but teacher accountability statements are not. Recall that alerting statements remind students of what they are responsible for doing. In contrast, accountability is the actual review of a student's work or request for student performance. Hence, alerting messages may be similar to time spent going over homework and review in that it provides students with clear expectations.

The data on page 2 of Table 6 are somewhat baffling. For example, it is not clear why correct answers and incorrect answers are both negatively associated with classroom climate while part right answers have a positive relation. Perhaps beyond a certain limit students resent frequent questioning (they see it as excessive). Although the correlations are quite small (and insignificant), students appear to like to be asked opinion questions and to call out responses. These data are reinforced by the somewhat stronger finding that students like to be asked self-reference questions.

Three or four interesting variables appear at the bottom of page 2 and the top of page 3. It can be seen that the frequency of no response or don't know responses followed by terminal feedback and incorrect responses followed by terminal feedback correlate negatively with affective mean scores. Presumably students do not like to have unsuccessful performance punctuated with terminal feedback. There is a very slight trend for students to prefer for their part right responses to receive terminal feedback. Perhaps this is due to the fact that part right responses leave the student psychologically more vulnerable (I should be able to answer this question, but I can't), therefore they are relieved when the teacher does not continue the interaction. However, broader differences between the two sets of teachers (those who obtained high and those who obtained low scores on the climate inventory) may control these isolated findings. Without the collection of more data under controlled circumstances it is very difficult to interpret the correlations that appear at the bottom of page 2 and the top of page 3. They are extremely interesting; however, and beg for experimental clarification. (However these findings describe behaviors that occur infrequently in the classroom.)

Still, these data are interesting because they suggest that a trade-off between cognitive and affective gains may be necessary in this phase of classroom activity. Alerting, enthusiasm, review time, and to some extent, time spent going over homework, do not necessarily have to be in conflict with cognitive gain. However, the fact that students prefer to receive terminal feedback after making part right responses seems to be in direct conflict with the finding that staying with students when they are part right is a more successful teaching strategy.

It is important to understand those process measures that may be conducive to both affective and cognitive growth and to distinguish those process measures that are uniquely related to growth in one of the areas. But at least equal consideration must be paid to process variables that relate to either an achievement or an affective outcome and that are negatively related to growth in the other area. Teacher feedback reactions to students in part right situations may be one such phenomenon.

Somewhat surprisingly, praise does not relate to affective (climate) scores. This may suggest that praise was used excessively or inappropriately (sugar-coating minimum student performance) by teachers in this sample.

These data, in consideration with the data reviewed above, raise strong suspicion about the utility of praise per se.

In general, the rest of the correlations that are reported in Table 6 are so low that they do not merit extended discussion and interpretation. As a set, however, there is nothing in the data to suggest that students react negatively toward being "forced" to seek the teacher out. Indeed, there is a slight tendency for students to feel more comfortable (at least to have higher affect scores) in environments where they seek the teacher out. In contrast, there is the consistent tendency for students to report lower affect scores in environments where the teacher seeks them out.

The generally low correlations that appear in Table 6 may be explained in part by the fact that the process measures selected for investigation in this study were chosen primarily for the purpose of explaining cognitive performance in mathematics. However, within these limits, the data "hint" that public events are probably more related to students' affective score than are private events. This is especially likely to be the case for public events that put a spotlight on the individual student. To learn more about those process measures that relate strongly toward positive affective expressions of students, it will be necessary to select classrooms for study on the basis of teachers' ability to produce high affect scores and the development of scales that focus upon affective variables. Table 7 spotlights the significant or near significant process measures that relate to affective means.

Teachers' mean pre-scores, post scores, raw gain scores, and residual gain scores are presented in Tables 8-11. These tables show that there is some relationship between pre-scores and the residual gain score (the operational definition of effectiveness in this study). However, it can also be seen that some teachers who started low did obtain good gains (both in a relative and an absolute sense). This can be seen more clearly in Table 12 where the rank order of the top nine and bottom nine teachers (those 18 teachers used in the analysis of variance process comparisons because of their stable patterns of effectiveness) are compared on pre-score and residual score.

Also in Table 11 it can be seen that those teachers who teach mathematics via group instruction tend to fall in the middle effectiveness range. No attempt here is made to describe the process milieux of these rooms, but it is interesting to see that teachers who utilize group instruction obtain average residual scores in comparison with fourth grade teachers using whole class instruction procedures.

Where Do Gains Come From?

Why is it that some teachers are relatively more effective in helping students to master mathematics skills (as measured by the total mathematics score on the Iowa Test of Basic Skills)? Is it because they get more gains from students who start the year with low, high, or middle achievement levels? If it were possible to identify systematic ways that effective teachers get their gains (e.g., by obtaining more from low achievement students than do other teachers), it would allow for powerful behavioral observation (e.g. focus on how teachers interact with low achievement students).

In an attempt to provide information about where teachers get their results, two different analyses were performed on a large number of third and fourth grade teachers. Residual gain scores for 104 third and fourth grade teachers were computed on the total mathematics score. These data were taken from fall testing 1972 and fall testing 1973.

Teachers, within grade levels, were assigned to one of three cells on the basis of their residual gain score (high, middle, or low). Similarly students within grade levels were assigned to one of three cells on the basis of their aptitude score on the Cognitive Abilities Test. In the first analysis, students were assigned to cells by dividing the entire population of third grade and fourth grade scores into three equal groups. Students in the top third were assigned to the high group and so forth. This analysis was called the <u>absolute analysis</u>.

The second analysis involved the assignment of students into cells on the basis of their aptitude rank within their mathematics class. In this analysis, students whose aptitude score was in the top third of their class were assigned to the high group (independent of how their aptitude compared to the total population of third graders or fourth graders). This analysis was called the <u>relative analysis</u>.

The results for both the relative and absolute analyses were similar: no interaction occurred between teacher competence and student aptitude. Two main effects were observed. Predictably, performance of high competence teachers exceeded the performance of middle and low competence teachers, and the performance of high aptitude students surpassed the performance of middle and low aptitude students (see Tables 13 and 14).

A similar analysis was performed on the pre-post data collected during the year of classroom observation. Since no recent aptitude score was available for students, it was necessary to do the analysis on the basis of students' pre-achievement scores. A residual mean score was computed for each teacher in the observational sample (on the basis of October 1974 and April 1975 test data). Teachers were assigned to high, middle, and low groups on the basis of their residual gain scores.

Students in each classroom were assigned to high, middle, and low cells on the basis of their pre-achievement level (measured in October 1974). Unsurprisingly, main effects were observed for both teacher competence and student achievement.

The results of an analysis over raw difference scores produced a small interaction effect between teaching competence and student achievement level. However, when the pre-score was used as a covariate in the analysis this interaction disappeared. The most accurate way to represent the relationship between teacher competence and student achievement is shown in Table 15. Clearly, teachers who are relatively effective (as do teachers who are relatively ineffective) show different effects on students. Teachers 24, 16, and 33 within the operational definition of this study are the most effective teachers. It can be seen that teacher 24 is especially helpful for low achievement students; whereas, teacher 16 is good with high achievement students and teacher 33 is notably poor with low achievement students. Hence, there appears to be no one way in which effective teachers get their gain. There is no evidence in this data set to suggest that one group of students benefits most from contact with high teachers or that a subgroup of students are disproportionately penalized from being in class with lows.

Discriminant Analysis

In order to examine the power of process data to discriminate relative effective and ineffective teachers, discriminant analyses were run across 22 of the 31 significant process-product relationships reported in Table 2. Some process variables were dropped from this analysis because the variable was not coded in some teachers' classrooms on one or more occasions. Hence if values were missing, the variable was dropped from the analysis.

The 24 variables that were included in the analyses appear in Table 16. The results are summarized in Tables 17-20. In Table 17, the results of the first discriminant analysis are reported and it can be seen that number of students in the classroom was the most important discriminator in the data set. Since teachers cannot effect (directly) the number of students they have and since teaching the whole class is not a relevant variable in some modes of instructions, two additional analyses were run. In the second analysis the variable number of students is omitted and in the third analysis this variable and time teaching the whole class are both omitted. A general summary of results appears in Table 20.

Clearly, the process data powerfully discriminate high and low teachers and it can be seen that the variable student created contact is the best discriminator. Consistently, in the rooms of high teachers there was more student initiated behavior and teaching time directed toward the whole class and less teacher warning and praise as well as fewer process questions. In high classrooms fewer questions of all types were asked (thus fewer incorrect and correct responses).

These data are important not because they describe mean differences between high and low classrooms (which they do), but because they describe process measures that consistently separate high and low classrooms. That is, high student initiation rates were a characteristic observed in all high classrooms.

Four high inference variables that appeared in Table 2 were not included in the discriminant analysis because of missing values. However, an examination of each teachers' mean score on climate, clarity, accountability and alerting show only clarity to be an excellent discriminator.

Most of the high teachers had higher clarity scores than lows. Climate, accountability, and alerting showed moderate discrimination (e.g. seven of the high teachers exceeded the mean performance of six lows on the climate rating), but there was an overlap between the two groups of classrooms (e.g. one low teacher got the highest climate rating).

Homework and Material Utilization

In general, coded information about homework assignments and the usage of materials in mathematics instruction yielded useful descriptive information but not information that highly differentiated the process in high and low classrooms. But a couple of findings do hint at importance differences. The highlights from these two codes follow.

Homework was assigned more frequently by highs than lows (48 percent versus 38 percent). However equal amounts of in class time was provided for completing homework. Furthermore high and low teachers in virtually all cases made the same assignment to all students.

Interestingly, information collected to describe homework assignments also provided useful information about the pace of classroom instruction (i.e. the speed with which the teacher is going through the textbook). It seems clear that highs are moving through the book more quickly than lows (recall all teachers were using the same textbook series). However quantifying these speed differences is difficult given the fact that some teachers skip around in the text rather than move directly from chapter to chapter. Subsequent work will provide more detail on this issue.

The concrete material checklist data suggest that teachers use materials infrequently. Although sixty-six percent of the teachers were observed to use materials during mathematics instruction, manipulative materials were utilized in but 19 percent of all observations that were completed (N = 265). Interestingly (despite the current Zeitgeist advocacy for the instructional use of materials), the data reveal a sizeable and negative rank correlation (-.71) between material usage and teacher residual means. Teachers who obtain high student performance as measured by the Iowa Achievement Test) do not make frequent use of materials. Extended presentation and discussion of the data describing material usage has been presented elsewhere (Grouws and Good, 1975).

Observers Comments

Roughly two weeks after the two observers had completed classroom coding, they were requested to complete a brief (one to two page) summary for each of the 41 classrooms. Observers were asked to comment upon teachers' general style. The intent here was to see if successful teachers were obvious: they weren't. In general descriptions of the highs were not particularly discriminating. That is, highs were not sharply differentiated from teachers who obtained average results. In contrast, all the brief descriptions of lows (with but one exception) indicated that learning problems existed in their classrooms. Comments about five of the lows appear below (drawn from observers' summaries).

> "After a practice assignment was given, students were allowed to move about freely with students working together. Student attention was very low with several students typically uncooperative. Teacher seemed confused and uneasy about the noise but did little about it. Discipline problems were frequent . . ."

> ". . . Teacher circles the room during written assignments. Students look around, talk, and loudly complain that they do not understand seatwork. Class disorganized . . . level of material appears too elementary . . ."

> "Explanations and examples were not understood by students judging from general student questions and responses. Most written work was taken from the text. Games were seldom used . . . when they were some students would say that they didn't want to participate. Work was checked in class by

exchanging papers. The teacher usually read answers aloud . . . occasionally asking a student to answer . . ."

"Students are noisy and constantly walk around the class. The teacher ignores the movement and the noise. Students generally have a 5-10 minute wait for help . . ."

". . . Teacher was dynamic . . . talked and moved rapidly. Class got loud during verbal practice sessions and teacher frequently turned out lights to quiet the class. Classroom was disorganized . . . teacher circulated during written assignments."

In general, the comments suggest that lows had general management and organization problems. Based upon our inspection of observers' notes it appears more difficult to spot effective teaching than it is to identify instances of ineffective teaching.

Discussion

The data demonstrate that it is possible to identify teachers who appear to make stable differences in students' learning of mathematics (as measured by the mathematics section of the Iowa Test of Basic Skills). Furthermore, the study suggests that it is also possible to identify teachers who create stable (and different) classroom climates year after year.*

This is especially interesting given the brief climate instrument that was used in this study. A more differentiated instrument (perhaps coupled with selective interviewing of students) could produce a sample of teachers who make the classroom experience more satisfying (without sacrificing learning gains). More information about the process that unfolds in such classrooms would be especially important data. Data about process antecedents of the affective and cognitive outcomes are both necessary if successful learning milieux are to be built.

The relationship between classroom climate and student achievement is not clear. In an initial study using all available teachers, the

*Performance stability is important as it represents a necessary condition for finding process-product relationships. But it must be emphasized that many teachers in the pre-observation sample showed wide fluctuation in achievement scores from year to year. This fact is an important restraint upon the use of process or product data in accountability plans.

correlation between climate and residual achievement was about zero. The correlation varied from subtest to subtest but all correlations were quite low. However, in the year of the observational study (2 years after the initial study) the obtained correlation between achievement (total math performance on the Iowa) and the same climate scale was .50.

We suspect two reasons account for this major difference. First, perhaps, by studying many teachers who were stable and relatively high in achievement we biased the sample systematically. The second and, we suspect, the more important reason is that during the year of observation students rated the teacher with particular reference to mathematics instruction. Earlier, students had rated the teacher and classroom instruction generally. Given that some students were shared for a part of the instructional day, this may have been a confusing task. In contrast, the observational study specifically targeted student reaction to instruction during mathematics.

We have reported a great deal of information about process differences in high and low achievement classrooms, but this study provides comparatively less (or at least weaker) information about process measures that relate to students' climate scores. However it should be remembered that the 18 teachers focused upon in the observational study had been selected because of their influence on students' achievement scores. Furthermore, process measures used to code classroom behavior had been selected primarily because it was believed that these measures would differentiate high and low achievement classrooms.

Still, the data identify several correlates of students' affective scores (climate). The most salient variable appears to be teacher enthusiasm.

Teachers who exhibit more enthusiasm receive higher climate scores from students. Students in this sample seemed to like alerting statements but disliked accountability statements. Time going over homework and review also seemed to be associated with higher affect scores. These results in combination seem to suggest that students enjoy structure and feedback but dislike being held accountable for their work. Review and homework discussion may be valued because of positive reasons (it communicates clear performance expectations) or because of negative reasons (review of mastered material frees students from the threat of failure on new tasks).

Higher affect scores are also associated with higher frequencies of teacher terminal behavior following students' part right responses. Students apparently prefer to end such interactions. But they apparently do not like to have their incorrect responses followed by teacher terminal behavior. That is, after failure (wrong answers or no response), students, as a group, prefer to continue the interaction . . . presumably in the attempt to save face. After making a part right response, students prefer termination. This may be because students see such interactions as sufficiently face-saving or because they don't like the threat of potential failure. Importantly, student achievement and satisfaction seem to be in conflict on this variable. Teachers' sustaining of interactions in which students initially provide part right answers appears to correlate positively to achievement gains but negatively to climate scores.

These data may be artifacts of other process differences, but they are interesting and, potentially, have major treatment implications that demand subsequent research. These data represent one of the few behavioral events where affective and cognitive growth appeared to be in direct conflict.

The process variables coded in this study identified a number of variables that separated high and low teachers. However rather than reviewing all of the variables here it seems more reasonable to discuss major clusters that appeared in the data.

Patterns of Effectiveness

Teaching effectiveness (as operationally defined in this study) appeared to be strongly associated with the following clusters: (1) student initiated behavior; (2) whole class instruction;* (3) general clarity of instruction, and availability of information as needed (process feedback in particular); (4) a non-evaluative and generally relaxed learning environment; (5) higher achievement expectations; (6) classrooms that were relatively free of major behavioral disorders.

Several different behavioral measures consistently demonstrated that high teachers were approached by students more than teachers in low classrooms. Presumably when students in high classrooms wanted information or evaluative input they felt free to approach the teacher. Even when the teacher dealt with the entire class in a public format, students in high rooms were able to participate by their own initiative. Students in these rooms asked the teacher more questions, called out more answers, and proportionately were asked more open questions (questions which students indicate they want to answer: they raise their hand, etc.).

In this context, student initiated behavior appears to make good sense. For example, students' call-out rates per hour are not excessive. Given a general population of middle class students, it appears appropriate to allow them to approach the teacher as they need help. Teachers who choose to rotate around the room will find the one or two students who are having

*This variable did not differentiate high and low classrooms as both utilized whole class instruction. However it was the instructional mode in high classrooms and these teachers obtained better results than teachers who taught mathematics via groups.

difficulty (and not approaching the teacher) but may delay several students who want teacher feedback. In a setting filled with students who possess at least minimal self-management skills, the general policy of allowing and encouraging students to approach the teacher is a good one. Teachers may profitably seek out the few students who don't come to them without developing systematic routines of circling the room.

A second general finding of these data was that students in high classes received instruction as a unit. They were given the same in-class assignments and identical homework assignments. However, students in low classes also were basically taught as a whole group. But low teachers also worked with individuals or groups of students much more frequently than did highs.

Perhaps if the variance of learners within a class is not too great, it makes more sense to gear instruction toward a particular mode (whole class or group: not both). Interestingly, those teachers who taught mathematics via groups in this particular study, fell into the middle effectiveness range. The data clearly suggest that teaching the class is not a poor or good strategy categorically. If the teacher possesses certain capabilities it may be an excellent strategy, if not, the whole class instructional mode will not work.

One of the necessary skills for effective whole class instruction is the ability to make clear presentations. Highs regularly exceeded lows in clarity scores. They generally introduced and explained material more clearly than did lows. Interestingly, in whole class settings highs asked more product questions and appeared to keep the "ball moving." However, when students did experience difficulty, highs were more likely to give process feedback than lows. In contrast, lows were more likely to ask process questions and less likely to give process feedback. It seems that highs did not focus upon process as a ritual, but rather, they used process responses when student responses indicated some error or misunderstanding.

The data demonstrate that high praise rates do not categorically enhance learning. Indeed, in this study, praise is negatively associated with both achievement and climate. Consistently, high teachers were found to praise less than low teachers. Interestingly, despite their high praise rates, lows were much less likely than highs to praise students when they approached them about their academic work. Presumably, low teachers prefer to go to students (rather than being approached by them), a strategy that proved to be ineffective in this study. High teachers were basically non-evaluative. They did not criticize or praise academic work as frequently as did low teachers. The evaluative stance of lows coupled with their high rates of approaching students may have interfered with learning progress as well as creating a "heavy" climate. High classrooms were regularly described more favorably by students, despite the fact that high teachers did not praise much.

Highs also appeared to demand more work and achievement from students. They assigned more homework and appear to move through the curriculum more briskly than did lows.

Low teachers seemed to have more frequent managerial problems than did high teachers. However, the data here are not as clear as for the five clusters described above. Several measures show little difference between high and low teachers (e.g. percent of students not involved in lesson). Suggestion of discipline problems stems from the fact that lows issue many more behavioral warnings and criticisms than do highs. This may be a comment upon differential teacher reaction to similar behavioral events (lows are more threatened by the same noise levels, etc.), but we suspect that there are more managerial problems in low classrooms. In part, we feel this way because students are often left sitting waiting for the teacher to come to them, and because they receive unclear and incomplete directions (as reflected in the high rates of teacher afforded contacts that were recorded in lows' classrooms and as reported in observers' summaries).

Operational Definitions

Surprisingly, we found that low teachers engage in more accountability and alerting activities than did highs. Two reasons explain this finding. First lows appear to "travel" the room more regularly than do highs and therefore are likely to check the work of more students in a direct fashion. (However, as we have pointed out above "checking" in rooms marked by frequent evaluative comment may interfere with student learning). However, an examination of the data, on page 1 of Table 1, shows that the means for both high and low teachers in this study is relatively high, implying that neither group engaged in much accountability or alerting within the operational definition of this study.

The second problem is the operational definition of accountability and alerting. In general, the definitions were taken from Kounin (1970) and applied to the project directly. That decision was a poor one. While the definitions make good sense for describing kindergarten or first grade classrooms, they do not for fourth grade classrooms. In brief, the definitions are too restrictive.

The basic issue in using the accountability and alerting codes was the number of students that teachers held accountable during a two minute instructional period (e.g. accountability was coded with the following checkpoints: from ½ to every student held accountable; from 25-50% held accountable; less than 25%; no accountability effort). Chorus questions and group hand-raising to teachers' questions or directions (Have you all finished? Did you get it right? Hold up your papers.) occurs less frequently than in kindergarten classrooms. A better operational definition would have dimensionalized more fully the 0 to 25% range and allowed for the coding of qualitative distinctions (appropriate or inappropriate). Still, the data are clear within the definitions of the study: lows engaged in more alerting and accountability actions than did highs.

More importantly the general pattern of results presented here must also be interpreted within the operational definition of effectiveness employed in this study: student performance on a standardized achievement test. Many alternative definitions of effectiveness exist and we hope that subsequent research will use multiple definitions in selecting teachers. Such research may identify differential process environments that are associated with different student outcomes. The definition we employed is but one way of looking at classroom progress. We feel that it is a valuable way to study classrooms, but we are also aware of the fact that any operational definition of effectiveness imposes restraints upon the investigation *per se* and the conclusions that are drawn.

Subsequent Analyses

The data reported here suggest that alerting and accountability may have nonlinear relationships with student achievement (there can be too much or too little. For example Brophy and Evertson (1974) and Kounin (1970) have provided data to suggest that ineffective teachers initiate too few alerting and accountability messages. The data in the present study suggests that, in some contexts, it may be possible for teachers to communicate too many accountability and alerting messages (especially if teachers are highly evaluative).

The data reported here describe only linear relationships between process measures and two product outcomes (student achievement, classroom climate). Subsequent work needs to center upon non-linear relations that may exist in the present data. Several findings reported in this study may be clarified and/or extended by subsequent data analysis activities. For example, praise, the frequency of teacher questions, and teacher afforded contact are probably variables that interfere with learning only if engaged in too frequently.

Furthermore, the data described here only report teacher behavior toward students generally. Student initiated work contact is analyzed as a student initiated work contact (whether a low achievement or high achievement student initiates the contact). Toward the end of the study, data were collected in some classrooms to describe interaction between the teacher and a few target students. When such data are analyzed, they may yield valuable clues that qualify and/or extend some of the data reported in this study.

However, the analysis of achievement patterns suggests that there are unlikely to be sharp, simple relationships that make more or less sense for students of varying achievement levels. Recall that absolute and relative analysis of residual achievement scores showed that highs <u>as a group</u> obtained a little better achievement from students of all levels. However, an examination of <u>individual</u> teacher's performance showed that teachers (with similar residual scores) get their gains from different achievement groups. Hence the sample of teachers who help low achievement students (or high achievement students) would be quite low (in this study).

Treatment Implications

The finding that individual teachers show varying effects on students from different achievement levels is an important finding. First, it provides explicit support for the contention that some teachers can work very effectively with low achievement students. But perhaps equally important it suggests that all teachers are (with the exception of teachers who are poorly organized and who experience high rates of misbehavior) probably effective with some students. Entering the classroom with the perspective of "what works for what type of student," is a totally different perspective than the one that typically motivates supervisors or researchers: what works and what doesn't. The pattern of achievement results (where teachers get their gain) suggests that the general question (what works and what doesn't) appears to be too broad to yield relevant decision making information. Hence, it would seem that subsequent naturalistic studies of teacher effectiveness could profitably study teachers who appear to be effective with certain types of students and to gear the observation to focus on previously identified teacher strengths and weaknesses. A conceptual model attempting to typologize students that elementary and secondary teachers must deal with has appeared elsewhere (Good and Power, in press).

To obtain more illuminating process-product relationships, it appears necessary to conceptualize better theories of the specific learning milieux that different types of learners used. This is not to discredit the important work that has taken place in teacher effectiveness research in the past five years. The field has provided useful information about general learning conditions that must be present (in some form and to some degree) in order to enhance the learning of young children. To go beyond these rough outlines of effectiveness it appears necessary to contextualize process-product studies.

The data reviewed in this study also appear to have practical import for the design and implementation of treatment programs. First, the data suggest that general effectiveness can be achieved in a variety of ways. If teachers are effective with some students prior to any intervention activity, it would seem necessary to design treatment programs that do not undermine effective patterns that exist in the classroom. Frequently, treatment programs focus upon the problems of particular learners and pay too little attention to the possibility that proposed action may help students at the expense of other students. It would seem important to collect process and product data prior to intervention efforts if one is to understand fully the impact of treatment programs on classroom life.

Further, we suspect that relatively simple treatment programs that focus upon one variable or dimension (e.g. have students approach the teacher) will do little to improve learning (and may even have a detrimental effect) if other dimensions are not also dealt with when necessary. That is, if one were designing a treatment program for teaching mathematics to middle class fourth graders, the six dimensions we presented earlier in the discussion would seem to be prime targets for experimental manipulation. However, manipulating the frequency of teacher initiated private contacts and ignoring evident managerial problems would appear to make little sense. Treatment programs need to deal with the total classroom process . . . just as classroom teachers must. And it would seem to make sense to tailor treatment programs to the context that the teacher works in and to design intervention strategies around existing teacher skills.

External Validity

In analyzing the data one general ground rule of statistical procedure was violated: more variables were compared than the number of teachers that were included in the study. The practical implication of this procedure falls upon the interpretation of probability levels. Namely, the probability levels reported for process measures cannot be interpreted in a direct straightforward way (i.e. the accuracy of the probability level is impossible to specify).

However, we preferred to collect data on a wide number of process variables and to link this data with achievement outcomes rather than to restrict the study to a few process measures. This decision was a good one because it allowed for the collection of a variety of measures that can be clarified in subsequent focused naturalistic studies or experimental paradigms.

Furthermore, probability levels under the best circumstances provide only a rough indication of population boundaries . . . an aroma of reality. The true test of process-product relationships lies in replication (do different studies draw the same conclusion) and experimental verification (if the process is manipulated in specified ways does the product vary in a predictable way?).

There are two ways to approach replication. First, is an analysis of the consistency of data collected in a particular study. In general, this is why we prefer to discuss clusters of behaviors rather than single behaviors. In this study, rather consistently, the clusters we have emphasized demonstrate strong internal consistency. High teachers individually show more or less of these behaviors than do low teachers. Furthermore, several behaviors subsumed within a cluster all fall in the same direction (e.g. on a variety of separate measures students in high rooms consistently initiate more behavior). Hence, the discriminant analysis and general consistency of behaviors within clusters provide strong evidence that the findings we emphasize are real associations rather than fluke, random relationships.

Yet another way to assess the general robustness of the data is to compare them with findings that others report. Presumably, chance findings or overly contextualized findings will drop out in such comparisons. There are several data sets that our results could be compared with, but the most interesting comparison source is the Texas Teacher Effectiveness Study (Brophy and Evertson, 1974). This is because both studies used the same basic coding system and studied a similar age group (Brophy and Evertson studied second and third grade classrooms: our study observed fourth grade classrooms). In particular, a comparison of our results with the process-product findings that Brophy and Evertson report for non title I schools represents a quasi replication.

In general, then, our data correspond to the linear process-product findings in high SES schools reported by Brophy and Evertson (1974). Probably the most basic correspondence is the pattern of results. In both data sets there are numerous weak correlations rather than a few big relations that seem to be of critical importance. The data in both studies suggest that successful teaching is based upon a large number of variables (that must be present to a minimum degree) rather than because of two or three critical factors.

Among the agreements in our data and the Brophy-Evertson high SES data are: calling on volunteers correlates positively with student achievement, student initiated questions correlate positively with achievement, teacher afforded work contact (going to students' seats) correlates negatively with student achievement, the positive relationship between process feedback and student achievement, the positive relationship of homework and achievement, and the negative value of using materials in instruction.

Both sets of findings suggest that it makes sense to allow students (who are capable of self-direction) to work semi-autonomously during seat work assignments, but also to allow these students the freedom to seek out the teacher when they need feedback. In public settings, both sets of findings suggest that within limits it is a reasonable strategy to call on volunteers, and that student initiated questions appear to be a sign of appropriate involvement.

However our data seem to go a little further in supporting the concept of student assertiveness and initiative than do the Brophy-Evertson data. For example student eagerness to respond (call-outs) correlated negatively in the Brophy-Evertson study but our data show a positive relationship. Perhaps this finding is because of age (older children are less likely to over respond) or because of subject matter (call-out is more appropriate in a focused subject like mathematics than other subjects).

For whatever reason the two studies provide conflicting results on this point. But again both data sets suggest that child initiative appears to be a generally good indicant of learning in middle SES classrooms. This form of initiative should not be confused with indirect teaching. . . a variable that has frequently been interpreted to mean: the less teachers talk the better. In general, we refer here to student initiative in seeking out the teacher during seat work and in seeking feedback during public discussion (by asking questions). These data say nothing about the frequency of student talk per se and in the context of the data presented in this study it appears that less frequent teacher questioning (and presumably less student talk) seems to be a more preferable instructional style.

There are other variables as well that appear to be in conflict when the present data and the Brophy-Evertson process-product are examined. For example clarity of teacher presentations seem to be an important variable in our study but it draws little support in the Texas data. In contrast one of the interesting variables in the Texas project (percent of correct answers) draws little confirmatory support in our data.

But as a set the findings appear to hang together reasonably well and provide a number of agreement "points" that can be directly translated into treatment research allowing the value of these correlational relationships to be directly tested. Obviously, points of conflict can also be studied in treatment research but initially they may need to be reexamined in other naturalistic settings as well.

Although the data do suggest a set of findings that are internally consistent and a set of clusters that draw solid replication support from the Texas Teacher Effectiveness project, the data should not be used for accountability purposes. The data are interesting and have clear treatment implications; however, until experimental work has been completed the process-product relationships reported here are appropriately viewed as hypotheses to be tested. We agree with Berliner (1975) that the study of teacher effectiveness is an extraordinarily complex task. Furthermore, it is a task that has started to produce useful information.

The cumulative contribution from a variety of research workers during the past few years has yielded major substantive and methodological insights that help in understanding why some classroom behaviors generally work. Continuation of this basic research with selective (theoretically and/or empirically based) experimental innovation, we feel, will eventually provide a rich data base that can be used to improve classroom instruction. Premature advocacy in the form of process accountability or general enthusiasm for a particular solution to teaching problems will needlessly interfere with the acquisition of dependable knowledge.

Table 1

Means and p Values on all Process Variables from an Analysis of Variance Across the Top and Bottom Nine Teachers

Variables	p Value	$\overline{X}$ High	$\overline{X}$ Low
Number of Students	.0001	26.70	21.34
Transition Time	.2191	2.45	3.14
Total Class Time	.8255	44.46	44.93
Time Teacher Taught "Whole" Class	.1001	40.47	35.83
Time Going Over Homework	.0656	4.98	8.19
Review Time	.6278	1.77	0.83
Development	.5519	7.25	5.71
Drill	.5777	14.35	16.78
Homework Practice	.1796	13.61	10.52
Homework Development	----	0.00	0.00
Organization*	.5317	1.47	1.55
Alerting*	.7636	2.42	2.38
Accountability*	.6547	2.16	2.02
Classroom Climate*	.0771	2.00	2.26
Managerial*	.6215	1.81	1.96
General Thrust of Homework*	.5129	1.33	1.69
Attention*	.2697	3.96	3.75
Clarity*	.0135	4.06	3.53
Enthusiasm*	.6732	3.56	3.37
Presentation*	.7660	3.19	3.26
Average Accountability*	.0424	3.46	3.15
Average Alerting*	.0350	3.90	3.59
% of Students "Probably" Involved*	.8294	6.89	6.29

*High inference variable

Variables	p Value	$\bar{X}$ High	$\bar{X}$ Low
% of Students "Definitely Not" Involved*	.8483	14.26	13.60
Student Asks Question	.5602	2.84	2.07
Teacher Asks St ent to Read	.9591	1.22	1.23
Discipline Type Question	.0656	0.11	0.35
Direct Question	.0113	14.07	28.27
Open Question	.6383	37.75	33.36
Student Calls Out Answer	.2098	3.96	1.20
Process Question	.0131	2.72	7.53
Product Question	.1490	44.34	55.00
Choice Question	.8390	2.50	2.25
Self Reference Question	.1538	0.85	0.27
Opinion Question	.6563	0.10	0.19
Correct Response	.0533	38.70	50.98
Partially Right Response	.5404	1.62	2.00
Wrong Response	.0017	5.49	10.61
"Don't Know" Response	.1862	0.31	0.61
"No Response"	.0058	1.37	3.26
Wrong Responses Followed by Teacher Criticism	.6022	6.65	10.57
Right Responses Followed by Teacher Praise	.5601	19.75	16.34
"No Responses" or "Don't Know" Responses Followed by Sustaining Feedback	.6238	2.85	3.57
"No Responses" or "Don't Know" Responses Followed by Terminal Feedback	.5284	1.88	2.16
Wrong Responses Followed by Terminal Feedback	.5413	2.23	2.49

Variables	p Value	$\bar{X}$ High	$\bar{X}$ Low
Wrong Responses Followed by Sustaining Feedback	.6716	3.57	4.40
Part Right Response Followed by Terminal Feedback	.1684	2.15	1.79
Part Right Response Followed by Sustaining Feedback	.2093	2.76	1.80
Praise	.0046	2.74	14.09
Affirm	.9398	34.57	34.20
Summarizes	.6516	0.73	1.18
No Teacher Feedback to Student's Answer	.5762	4.27	6.44
Negate Wrong	.0088	1.51	3.29
Criticism	.8581	0.49	0.45
Process Feedback	.2583	1.73	0.98
Gives Answer	.6672	1.73	2.07
Asks Another Student	.2268	5.03	7.00
Another Student Calls Out Answer	.5239	0.15	0.23
Repeats Question	.0295	1.39	2.78
Gives Clue	.9747	2.18	2.20
Asks New Question	.6818	5.33	4.73
Expands Student's Response	.6221	0.44	0.28
Student Initiated Work Related Contact--Teacher Praises	.1556	1.03	0.45
Student Initiated Work Related Contact--Teacher Gives Process Type Feedback	.0654	4.41	1.56
Student Initiated Work Related Contact--Teacher Gives Feedback	.0004	17.65	9.30
Student Initiated Work Related Contact--Teacher Criticizes	.9438	0.28	0.26
Student Initiated Work Related Contact--Type Teacher Feedback Unknown	.1260	0.06	0.23

Variables	p Value	$\bar{X}$ High	$\bar{X}$ Low
Student Initiated Procedure Related Contact--Teacher Praises	.3163	0.20	0.00
Student Initiated Procedure Related Contact--Teacher Gives Feedback	.8059	1.66	1.55
Student Initiated Procedure Related Contact--Teacher Criticizes	.8122	0.05	0.06
Teacher Initiated Work Related Contact--Teacher Gives Praise**	.2308	0.39	1.54
Teacher Initiated Work Related Contact--Teacher Gives Process Feedback	.5303	0.37	0.49
Teacher Initiated Work Related Contact--Teacher Gives Feedback	.1649	1.99	3.29
Teacher Initiated Work Related Contact--Teacher Criticizes	.2997	0.24	0.42
Teacher Initiated Work Related Contact--Type Teacher Feedback Unknown	.1072	0.02	0.24
Teacher Initiated Behavior Related Contact--Teacher Gives Process Feedback	.5177	2.13	1.74
Teacher Initiated Behavior Related Contact--Teacher Praises	.6364	0.05	0.07
Teacher Initiated Behavior Related Contact--Teacher Warns Student	.0081	1.75	3.37
Teacher Initiated Behavior Related Contact--Teacher Criticizes Student	.0548	0.30	0.67
Total Response Opportunities	.5860	59.82	66.49
Total Teacher Initiated Work Related Contacts	.0383	3.01	5.98
Total Teacher Initiated Behavior Related Contacts	.0853	4.22	5.86
Total Teacher Initiated Contacts	.0129	7.23	11.83
Total Student Initiated Work Related Contacts	.0004	23.44	11.80

**The term 'teacher initiated' is synonymous with the term 'teacher afforded' that is used in the appendix section that contains coding definitions.

Variables	p Value	$\bar{X}$ High	$\bar{X}$ Low
Total Student Initiated Procedure Related Contacts	.5748	1.91	1.61
Total Student Initiated Contacts	.0003	25.35	13.41
Total Dyadic Contacts (Student Initiated, Teacher Initiated and Response Opportunities)	.9521	92.41	91.73
Direct Questions / Direct Plus Open Questions	.2614	33.63	40.21
Direct Questions / Response Opportunities	.1089	28.13	36.54
Open Questions / Response Opportunities	.5404	58.06	53.81
Call Outs / Response Opportunities	.8105	3.61	3.28
Student Initiated Work Related Contacts / Total Student Initiated Contacts	.5734	89.22	85.79
Teacher Initiated Work Related Contacts / Total Teacher Initiated Contacts	.2759	33.60	41.31
Total Teacher Initiated Contacts / Total Student Initiated Contacts	.0058	54.10	116.41
Process Questions / Total Questions	.0518	7.44	14.56
Choice Questions / Total Questions	.6496	2.19	3.10
Opinion Questions / Total Questions	.6126	0.05	0.15
Product Questions / Total Questions	.1950	94.93	85.70
Correct Responses / Total Response	.0051	82.80	76.17
Wrong Response / Wrong Response Plus No Response	.6032	76.99	74.42
Don't Know / Don't Know Plus No Response	.6347	16.09	19.98

Variables	p Value	$\bar{X}$ High	$\bar{X}$ Low
% of Responses Teacher Gave No Feedback	.5760	6.54	8.61
Process Feedback / Response Opportunities	.1859	5.89	1.90
Process Feedback / Product Feedback	.2273	7.15	3.81
Expands Feedback / Total Feedback	.7938	12.68	12.17
Process Feedback in Student Initiated Work Related Contacts / Total Student Initiated Work Related Contact	.5468	15.80	12.80
Process Feedback in Teacher Initiated Work Related Contacts / Total Teacher Initiated Work Related Contacts	.9246	14.25	13.65
Total Process Feedback	.1005	6.51	3.04

Table 2

Significant or Near Significant Process Variables from an Analysis of Variance across the Top and Bottom Nine Teachers

Variables	p Value	$\bar{X}$ High	$\bar{X}$ Low
Number of Students	.0001	26.70	21.34
Time Teacher Taught "Whole" Class	.1001	40.47	35.83
Time Going Over Homework	.0656	4.98	8.19
Classroom Climate	.0771	2.00	2.26
Clarity	.0135	4.06	3.53
Average Accountability	.0424	3.46	3.15
Average Alerting	.0350	3.90	3.59
Discipline Question	.0656	0.11	0.35
Direct Question	.0113	14.07	28.26
Process Question	.0131	2.72	7.53
Correct Response	.0533	38.70	50.98
Wrong Response	.0017	5.39	11.39
No Response	.0058	1.37	3.26
Student Response Followed by Teacher Praise	.0046	2.74	14.09
Negates Wrong	.0088	1.51	3.29
Repeats Question	.0295	1.39	2.78
Student Initiated Work Related Contact; Teacher Gives Process Feedback	.0654	4.41	1.56
Student Initiated Work Related Contact: Teacher Gives Feedback	.0004	17.65	9.30
Teacher Initiated Work Related Contact; Type Feedback Unknown	.1072	0.02	0.24
Teacher Initiated Behavior Related Contact; Teacher Gives Warning	.0081	1.75	3.37

Variables	p Value	$\bar{X}$ High	$\bar{X}$ Low
Teacher Initiated Behavior Related Contact; Teacher Gives Criticism	.0548	0.30	0.67
Total Teacher Initiated Work Related Contacts	.0383	3.01	5.96
Total Teacher Initiated Behavior Related Contacts	.0853	4.22	5.85
Total Teacher Initiated Contacts	.0129	7.23	11.83
Total Student Initiated Work Related Contacts	.0004	23.44	11.80
Total Student Initiated Contacts (Work and Procedural)	.0003	25.35	13.41
Direct Questions / Total Response Opportunities	.1089	28.13	36.54
Total Teacher Initiated Contacts / Total Student Initiated Contacts	.0058	54.10	116.41
Process Questions / Total Questions	.0518	7.44	14.56
Correct Responses / Total Responses	.0051	82.80	76.17
Total Process Feedback	.1005	6.51	3.04

Table 3

Correlations of Behavioral Measures with Teachers' Residual Means

Variables	Correlation	p Values
Affective Mean Score	.50	.0038
Number of Students	.43	.0115
Transition Time	.11	.5538
Total Class Time	.18	.3244
Time Teacher Taught "Whole" Class	.09	.6389
Time Going Over Homework	.02	.9182
Review Time	.29	.1009
Development	-.13	.5076
Drill	.11	.5655
Homework Practice	.08	.6566
Homework Development	--	--
Organization	-.02	.9312
Alerting	.04	.8009
Accountability	-.04	.8394
Classroom Climate	-.28	.1159
Managerial	<.00	.9765
General Thrust of Homework	.02	.9463
Attention	-.09	.6386
Clarity	.22	.2527
Enthusiasm	.02	.9078
Presentation	-.38	.0459
Average Accountability	.22	.2221
Average Alerting	.24	.1728

Variables	Correlation	p Values
% of Students "Probably" Involved	-.07	.6862
% of Students "Definitely Not" Involved	.01	.9401
Student Asks Question	.09	.6094
Teacher Asks Student to Read	.14	.5550
Discipline Question	-.30	.0852
Direct Question	-.08	.6592
Open Question	.17	.6426
Student Calls Out Answer	.32	.0682
Process Question	-.15	.6040
Product Question	-.02	.8967
Choice Question	.15	.5921
Self Reference Question	.20	.2561
Opinion Question	.05	.7612
Correct Response	-.03	.8594
Partially Right Response	-.13	.5109
Wrong Response	-.20	.2659
"Don't Know" Response	-.22	.2080
"No Response"	-.19	.2838
Wrong Responses Followed by Teacher Criticism	-.15	.6151
Right Responses Followed by Teacher Praise	.09	.6256
"No Response" or "Don't Know" Responses Followed by Sustaining Feedback	-.21	.2463
"No Response" or "Don't Know" Responses Followed by Terminal Feedback	-.22	.2317

Variables	Correlation	p Values
Wrong Responses Followed by Terminal Feedback	-.19	.2938
Wrong Responses Followed by Sustaining Feedback	-.10	.5676
"Part Right" Response Followed by Terminal Feedback	.11	.5820
"Part Right" Response Followed by Sustaining Feedback	.30	.1269
Praise	-.18	.6714
Affirm	.15	.5744
Summarizes	-.13	.5062
No Teacher Feedback to Student's Response	.03	.8418
Negate Wrong	-.04	.8049
Criticism	.02	.9167
Process Feedback	.27	.1279
Gives Answer	-.02	.9082
Asks Another Student	-.07	.7038
Another Student Calls Out Answer	-.15	.5929
Repeats Question	-.07	.6882
Gives Clue	.03	.8722
Asks New Question	.12	.5161
Expands Student's Response	.02	.9265
Student Initiated Work Related Contact--Teacher Praises	.28	.1100
Student Initiated Work Related Contact--Teacher Gives Process Type Feedback	.14	.5647
Student Initiated Work Related Contact--Teacher Gives Feedback	.37	.0312

Variables	Correlation	p Values
Student Initiated Work Related Contact--Teacher Criticizes	-.11	.5529
Student Initiated Work Related Contact--Type Teacher Feedback Unknown	-.21	.2424
Student Initiated Procedure Related Contact--Teacher Praises	.18	.3050
Student Initiated Procedure Related Contact--Teacher Gives Feedback	-.05	.7664
Student Initiated Procedure Related Contact--Teacher Criticizes	-.10	.5754
Teacher Initiated Work Related Contact--Teacher Gives Praise	-.14	.5627
Teacher Initiated Work Related Contact--Teacher Gives Process Feedback	-.29	.1037
Teacher Initiated Work Related Contact--Teacher Gives Feedback	-.25	.1593
Teacher Initiated Work Related Contact--Teacher Criticizes	-.19	.3016
Teacher Initiated Work Related Contact--Type Teacher Feedback Unknown	-.28	.1134
Teacher Initiated Behavior Related Contact--Teacher Gives Process Feedback	.02	.8865
Teacher Initiated Behavior Related Contact--Teacher Praises	.05	.7943
Teacher Initiated Behavior Related Contact--Teacher Warns Student	-.30	.0813
Teacher Initiated Behavior Related Contact--Teacher Criticizes Student	-.05	.7684
Total Response Opportunities	.14	.5555
Total Teacher Initiated Work Related Contacts	-.33	.0603

Variables	Correlation	p Values
Total Teacher Initiated Behavior Related Contacts	-.21	.2301
Total Teacher Initiated Contacts	-.33	.0608
Total Student Initiated Work Related Contacts	.37	.0344
Total Student Initiated Procedure Related Contacts	-.03	.8795
Total Student Initiated Contacts	.35	.0427
Total Dyadic Contacts (Student Initiated, Teacher Initiated, and Response Opportunities)	.17	.6687
Direct Question / Direct Plus Open Question	.05	.7617
Direct Question / Response Opportunities	.02	.9181
Open Questions / Response Opportunities	-.09	.6254
Call Outs / Response Opportunities	<.00	.9828
Student Initiated Work Related Contacts / Total Student Initiated Contacts	.01	.9687
Teacher Initiated Work Related Contacts / Total Teacher Initiated Contacts	-.26	.1380
Total Teacher Initiated Contacts / Total Student Initiated Contacts	-.34	.0490
Process Questions / Total Questions	-.19	.2830
Choice Questions / Total Questions	-.25	.1506
Opinion Questions / Total Questions	-.03	.8447
Product Questions / Total Questions	-.10	.6044

Variables	Correlation	p Values
Correct Responses / Total Responses	.25	.1531
Wrong Responses / Wrong Responses Plus "No Response"	.19	.2788
"Don't Know" / "Don't Know" Plus "No Response"	-.16	.6130
% of Responses Teacher Gave No Feedback	-.07	.7158
Process Feedback / Response Opportunities	.24	.1727
Process Feedback / Product Feedback	.25	.1520
Expands Feedback / Total Feedback	-.15	.5760
Process Feedback in Student Initiated Work Related Contacts / Total Student Initiated Work Related Contacts	-.19	.3017
Process Feedback in Teacher Initiated Work Related Contacts / Total Teacher Initiated Work Related Contacts	-.20	.2539
Total Process Feedback	.16	.6117

Table 4

Significant and Near Significant Correlations of Behavioral Measure and Teachers' Mean Residual Scores

Variables	Correlation	p Values
Affective Mean Score	.50	.0038
Total Number of Students	.43	.0115
Teacher Reviews Lesson	.29	.1009
General Classroom Climate	-.28	.1159
Teacher Presentation	-.38	.0459
Teacher Asks Discipline Question	-.30	.0852
Student Calls Out Answer	.32	.0652
Student Initiated Work Related Contact; Teacher Praises	.28	.1100
Student Initiated Work Related Contact; Teacher Gives Feedback	.37	.0312
Teacher Initiated Work Related Contact; Teacher Gives Process Feedback	-.29	.1034
Teacher Initiated Work Related Contact; Type of Teacher Feed-back Unknown	-.28	.1134
Teacher Initiated Behavioral Related Contact; Teacher Gives Warning	-.30	.0813
Total Teacher Initiated Work Related Contacts	-.33	.0603
Total Teacher Initiated Contacts (Work and Behavioral)	-.33	.0608
Total Student Initiated Contacts (Work and Procedural)	.35	.0427
Total Teacher Initiated Contacts Total Student Initiated Contacts	-.34	.0490
Total Student Initiated Work Related Contacts	.37	.0344

Table 5

Significant or Near Significant Process Variables from an Analysis of Variance Across the Top and Bottom Three Teachers

Variables	p Values	$\bar{X}$ High	$\bar{X}$ Low
Number of Students	.0027	27.90	20.33
Time Teacher Taught "Whole" Class	.0284	37.90	25.50
Classroom Climate	.0082	1.85	2.53
Average Accountability	.0190	3.47	2.78
Discipline Question	.0317	0.00	0.59
Process Question	.0511	1.89	5.31
No Response from Student	.0353	0.74	2.06
"No Response" or "Don't Know" Response Followed by Sustaining Feedback	.0988	1.88	4.54
Student Response Followed by Teacher Praise	.0382	1.92	7.75
Repeats Question	.0876	0.48	1.21
Gives Clue	.0819	0.78	1.67
Teacher Expands Student's Response	.0683	0.00	0.25
Student Initiated Work Related Contact; Teacher Gives Feedback	.0003	26.08	9.08
Student Initiated Procedural Contact; Teacher Gives Feedback	.0719	1.08	2.67
Teacher Initiated Work Related Contact; Teacher Gives Praise	.0706	0.19	0.89
Teacher Initiated Work Related Contact; Teacher Gives Feedback	.0024	0.95	5.86
Teacher Initiated Work Related Contact; Teacher Criticizes Work	.0205	0.00	0.46
Teacher Gives Behavioral Warning	.0023	1.08	4.93
Teacher Gives Behavioral Criticism	.0028	0.00	1.14

Variables	p Value	$\bar{X}$ High	$\bar{X}$ Low
Total Teacher Initiated Work Related Contacts	.0016	1.52	8.41
Total Teacher Initiated Behavioral Contacts	.0022	2.65	8.34
Total Teacher Initiated Contacts (Work and Behavior)	.0002	4.17	16.75
Student Initiated Work Related Contacts	.0012	30.26	13.03
Student Initiated Procedural Contacts	.0431	1.08	2.88
Total Student Initiated Contacts (Work and Procedural)	.0030	31.34	15.91
Direct Questions / Direct Questions Plus Open Question	.0121	21.44	48.63
Direct Questions / Total Response Opportunities	.0210	19.05	40.60
Open Questions / Total Response Opportunities	.0172	72.88	47.29
Student Initiated Work Contacts / Total Student Initiated Contacts (Work and Procedural)	.0064	96.22	81.17
Total Teacher Initiated Contacts / Total Student Initiated Contacts	.0004	26.02	122.30
Teacher Process Feedback / Teacher Product Feedback	.0965	6.44	2.15

Table 6

Correlations of Behavioral Measures with Teachers' Affective Means

Variables	Correlation	p Values
Number of Students	.15	.5862
Transition Time	.26	.1435
Total Class Time	.35	.0478
Time Teacher Taught "Whole" Class	.16	.6169
Time Going Over Homework	.34	.0537
Review Time	.42	.0161
Development	-.08	.6659
Drill	-.14	.5617
Homework Practice	.03	.8707
Homework Development	--	--
Organization	-.07	.6945
Alerting	.19	.3041
Accountability	-.02	.8939
Classroom Climate	-.29	.1016
Managerial	-.02	.9198
General Thrust of Homework	.17	.5300
Attention	.23	.2530
Clarity	.19	.6592
Enthusiasm	.47	.0125
Presentation	-.17	.6018
Average Accountability	.04	.8455
Average Alerting	.38	.0314

Variables	Correlation	p Values
% of Students "Probably" Involved	-.09	.6416
% of Students "Definitely Not" Involved	-.02	.9132
Student Asks Question	<-.00	.9805
Teacher Asks Student to Read	.15	.5964
Discipline Question	-.21	.2390
Direct Question	-.19	.3059
Open Question	-.11	.5419
Student Calls Out Answer	.17	.6431
Process Question	-.03	.8523
Product Question	-.27	.1278
Choice Question	-.02	.9038
Self Reference Question	.28	.1151
Opinion Question	.16	.6086
Correct Response	-.29	.1054
Partially Right Response	.10	.5881
Wrong Response	-.30	.0900
"Don't Know" Response	.01	.9199
"No Response"	-.09	.6304
Wrong Responses Followed by Teacher Criticism	.12	.6953
Right Responses Followed by Teacher Praise	-.10	.6123
"No Response" or "Don't Know" Responses Followed by Sustaining Feedback	-.03	.8815
"No Response" or "Don't Know" Responses Followed by Terminal Feedback	-.43	.0149

Variables	Correlation	p Values
Wrong Responses Followed by Terminal Feedback	-.41	.0208
Wrong Responses Followed by Sustaining Feedback	-.09	.6480
"Part Right" Response Followed by Terminal Feedback	.25	.2191
"Part Right" Response Followed by Sustaining Feedback	-.03	.8963
Praise	-.37	.0372
Affirm	.04	.8219
Summarizes	-.16	.6179
No Teacher Feedback to Student's Response	-.20	.2674
Negate Wrong	<-.00	.9941
Criticism	.24	.1899
Process Feedback	.10	.5799
Gives Answer	-.03	.8595
Asks Another Student	-.10	.5918
Another Student Calls Out Answer	-.18	.6642
Repeats Question	-.26	.1512
Gives Clue	-.14	.5568
Asks New Question	.13	.5284
Expands Student's Response	-.06	.7542
Student Initiated Work Related Contact--Teacher Praises	.19	.2947
Student Initiated Work Related Contact--Teacher Gives Process Type Feedback	.07	.7168
Student Initiated Work Related Contact--Teacher Gives Feedback	.26	.1548

Variables	Correlation	p Values
Student Initiated Work Related Contact--Teacher Criticizes	-.10	.5960
Student Initiated Work Related Contact--Type Teacher Feedback Unknown	-.39	.0263
Student Initiated Procedure Related Contact--Teacher Praises	.14	.5339
Student Initiated Procedure Related Contact--Teacher Gives Feedback	-.18	.3258
Student Initiated Procedure Related Contact--Teacher Criticizes	-.14	.5429
Teacher Initiated Work Related Contact--Teacher Gives Praise	-.29	.0982
Teacher Initiated Work Related Contact--Teacher Gives Process Feedback	-.27	.1385
Teacher Initiated Work Related Contact--Teacher Gives Feedback	-.09	.6363
Teacher Initiated Work Related Contact--Teacher Criticizes	-.05	.7831
Teacher Initiated Work Related Contact--Type Teacher Feedback Unknown	-.16	.5958
Teacher Initiated Behavior Related Contact--Teacher Gives Process Feedback	.19	.6707
Teacher Initiated Behavior Related Contact--Teacher Praises	-.25	.1618
Teacher Initiated Behavior Related Contact--Teacher Warns Student	-.07	.7132
Teacher Initiated Behavior Related Contact--Teacher Criticizes Student	-.06	.7357
Total Response Opportunities	-.15	.5680

Variables	Correlation	p Values
Total Teacher Initiated Work Related Contacts	-.25	.1643
Total Teacher Initiated Behavior Related Contacts	-.02	.8891
Total Teacher Initiated Contacts	-.15	.5830
Total Student Initiated Work Related Contacts	.23	.1989
Total Student Initiated Procedure Related Contacts	-.16	.6153
Total Student Initiated Contacts	.20	.2711
Total Dyadic Contacts (Student Initiated, Teacher Initiated, and Response Opportunities)	-.12	.5299
Direct Questions / Direct Plus Open Question	-.02	.8986
Direct Questions / Response Opportunities	-.04	.8432
Open Questions / Response Opportunities	.05	.7995
Call Outs / Response Opportunities	-.19	.2896
Student Initiated Work Related Contacts / Total Student Initiated Contacts	.14	.5564
Teacher Initiated Work Related Contacts / Total Teacher Initiated Contacts	-.08	.6761
Total Teacher Initiated Contacts / Total Student Initiated Contacts	-.15	.5983
Process Questions / Total Questions	-.02	.9111
Choice Questions / Total Questions	-.17	.6587
Opinion Questions / Total Questions	.04	.8189

Variables	Correlation	p Values
Product Questions / Total Questions	.10	.5992
Correct Responses / Total Responses	-.02	.9277
Wrong Responses / Wrong Responses Plus "No Response"	-.14	.5528
"Don't Know" / "Don't Know" Plus "No Response"	.02	.9194
% of Responses Teacher Gave No Feedback	-.21	.2436
Process Feedback / Response Opportunities	.07	.6953
Process Feedback / Product Feedback	.11	.5534
Expands Feedback / Total Feedback	<.00	.9733
Process Feedback in Student Initiated Work Related Contacts / Total Student Initiated Work Related Contacts	.07	.6948
*Process Feedback in Teacher Initiated Work Related Contacts / Total Teacher Initiated Work Related Contacts	-.17	.6578
Total Process Feedback	.05	.7875

Table 7

Significant and Near Significant Correlations of Behavioral Measures with Teachers' Affective Means

Variables	Correlation	p Values
Residual Mean Score	.50	.0038
Time Teacher Taught "Whole" Class	.35	.0478
Time Going Over Homework	.34	.0537
Review Time	.42	.0161
General Classroom Climate	-.29	.1016
Teacher Enthusiasm During Presentation	.47	.0125
Average Alerting	.38	.0314
Teacher Asks Self Reference Question	.28	.1151
Correct Student Response	-.29	.1054
No Response or "Don't Know" Responses Followed by Terminal Feedback	-.43	.0149
Wrong Responses Followed by Terminal Feedback	-.41	.0208
Student's Response Followed by Teacher Praise	-.37	.0372
Student Initiated Work Related Contact. Type of Teacher Feedback Unknown	-.39	.0263
Teacher Initiated Work Related Contact. Teacher Praises	-.29	.0982

Table 8*

Raw Gain in Each Class on Iowa Mathematics Subtests

Teacher #	M1	M2	M
01	.50	.29	.40
02	.72	.48	.57
03	.88	.77	.81
04	.76	.73	.73
05	.93	.11	52
06	.63	.56	.59
07	.44	-.14	.14
08	.25	.41	.33
09	.36	.50	.53
10	.77	.46	.64
11	.80	.47	.64
12	.60	.35	.47
13	.83	.41	.62
14	.64	.37	.55
15	.74	.66	.71
16	.96	.83	.89
17	.23	.12	.19
18	.76	.75	.75
19	.90	.54	.73
20	.34	.61	.48
21	.35	.15	.21
22	.86	.77	.81
23	.71	.90	.80
24	.80	.89	.86
25	NO	PRE	MEAN
26	.72	.50	.60
27	.76	.74	.75
28	1.01	1.36	1.18
29	.43	.47	.44
30	.83	.37	.60
31	.43	.49	.46
32	.99	.39	.68

*The data was not used in some classrooms to compute residuals when it was found that a class contained third and fifth grade students as well as fourth graders.

Teacher #	M1	M2	M
33	.73	.81	.79
34	NO	PRE	MEAN
35	NO	PRE	MEAN
36	.29	.35	.31
37	.15	.69	.40
38	.46	.31	.37
39	.80	.48	.62
40	.52	-.14	.19
41	.63	.60	.63

Table 9

Pre-Post Correlations by Teachers for Mathematics Achievement

Teacher	M_1	M_2	M
1	.73	.57	.73
2	.44	.66	.68
3	.63	.59	.68
4	.86	.72	.86
5	.78	.80	.90
6	.22	.06	.75
7	.21	.30	.51
8	.54	.57	.61
9	.77	.79	.90
10	.60	.48	.67
11	.73	.48	.71
12	.85	.80	.92
13	.79	.68	.85
14	.86	.64	.87
15	.41	.45	.60
16	.89	.84	.90
17	.76	.78	.88
18	.90	.82	.90
19	.77	.50	.72
20	.72	.82	.87
21	.34	.29	.49
22	.77	.43	.73
23	.53	.29	.33

Teacher	M_1	M_2	M
24	.59	.60	.77
25	.00	.00	.00
26	.50	.50	.63
27	.79	.61	.72
28	.28	.49	.39
29	.85	.79	.89
30	.87	.76	.90
31	.80	.70	.87
32	.54	.73	.59
33	.91	.52	.85
34	.00	.00	.00
35	.00	.00	.00
36	.41	.32	.40
37	.00	.00	.00
38	.68	.66	.72
39	.75	.56	.80
40	.49	.20	.53
41	.99	.63	.84

Table 10

Pre, Post, and Residual Means for Teachers During Time Period when Behavioral Data were Collected

Teacher	Pre Mean			Post Mean			Residual		
	M_1	M_2	M	M_1	M_2	M	M_1	M_2	M
01	4.91	4.71	4.80	5.41	5.00	5.20	-.068	-.110	-.126
02	3.75	3.64	3.70	4.47	4.12	4.27	-.106	-.135	-.034
03	4.65	4.63	4.65	5.53	5.40	5.46	.143	.359	.197
04	4.27	3.80	4.04	5.03	4.53	4.77	.036	.058	.060
05	3.78	4.02	3.90	4.71	4.13	4.42	.042	-.513	-.195
06	4.81	4.34	4.58	5.44	4.90	5.17			
07	3.56	3.42	3.49	4.00	3.28	3.63	-.341	-.835	-.511
08	4.73	4.42	4.57	4.98	4.83	4.90	-.270	-.027	-.152
09	4.90	4.62	4.76	5.26	5.12	5.29	-.342	-.088	-.111
10	4.71	4.77	4.74	5.48	5.23	5.38	.108	.021	.031
11	3.88	3.50	3.69	4.68	3.97	4.33	.053	-.190	.001
12	4.94	4.78	4.86	5.54	5.13	5.33	.012	-.051	-.074
13	4.75	4.54	4.64	5.58	4.95	5.26	.227	.007	.091
14	4.55	4.30	4.42	5.19	4.67	4.97	.008	-.130	-.050
15	5.08	4.80	4.93	5.82	5.46	5.64			
16	4.34	4.28	4.31	5.30	5.11	5.20	.260	.334	.295
17	4.98	4.59	4.78	5.21	4.71	4.97	-.295	-.265	-.313
18	4.66	4.46	4.57	5.42	5.21	5.32	.124	.288	.171
19	4.58	4.55	4.57	5.48	5.09	5.30	.277	.130	.190
20	4.52	4.27	4.39	4.86	4.88	4.87	-.331	.098	-.118
21	3.80	3.75	3.79	4.15	3.90	4.00	-.317	-.335	-.298
22	4.28	3.97	4.13	5.14	4.74	4.94	.214	.132	.189
23	3.39	3.32	3.31	4.10	4.13	4.11	-.156	-.129	.084
24	5.35	4.96	5.15	6.15	5.85	6.01	.318	.569	.379
15				5.69	5.60	5.64			
26	4.00	3.99	4.02	4.72	4.49	4.62	.086	-.043	.061
27	4.40	4.13	4.27	5.16	4.87	5.02	.096	.267	.202
28	5.00	4.45	4.73	6.01	5.81	5.91			
29	4.67	4.55	4.61	5.10	5.02	5.05	-.146	.080	-.075
30	4.32	4.53	4.12	5.15	4.90	5.02	.217	-.030	.088
31	3.97	3.65	3.81	4.40	4.15	4.27	-.398	-.210	-.244

Teacher	Pre Mean			Post Mean			Residual		
	M_1	M_2	M	M_1	M_2	M	M_1	M_2	M
32	5.37	5.52	5.43	6.36	5.90	6.11			
33	5.41	5.11	5.25	6.14	5.92	6.04	.209	.390	.232
34				3.41	3.54	3.47			
35				4.10	4.11	4.11			
36	3.59	3.39	3.49	3.88	3.74	3.80	-.410	-.328	-.292
37	4.65	3.92	4.30	4.80	4.61	4.70			
38	4.24	3.89	4.08	4.70	4.20	4.45	-.228	-.266	-.225
39	5.43	5.11	5.28	6.23	5.59	5.90	.316	.201	.154
40	3.40	3.73	3.57	3.92	3.59	3.76	-.266	-.680	-.399
41	4.51	3.99	4.23	5.14	4.59	4.86	-.016	.071	.065

Table 11

RESIDUAL MEAN SCORES (M), AFFECT MEANS AND RANKS

Teacher	M	M R	x Affect	R
24*	.379	1	33.6	9
16*	.295	2	33.0	11
33*	.231	3	33.9	6
27*	.201	4	34.7	4
3*	.197	5	32.3	16
19*	.190	6	35.4	1
22*	.188	7	32.9	12
18*	.171	8	32.0	18
39*	.154	9	35.0	2
13	.091	10	31.7	20
30**	.088	11	34.8	3
23	.084	12	33.0	11
41	.065	13	30.0	27
26**	.061	14	30.4	26
4**	.060	15	32.8	13
10	.030	16	28.4	30
11**	.001	17	28.4	30
2**	-.034	18	33.8	7
14	-.050	19	30.5	25
12**	-.074	20	30.9	22
29**	-.075	21	32.3	16
9*	-.111	22	33.1	10
20*	-.118	23	----	
1*	-.126	24	32.1	17
8*	-.152	25	32.1	17
5	-.195	26	30.7	24
38	-.224	27	32.0	18
31*	-.243	28	27.0	31
36	-.292	29	31.8	19
21*	-.298	30	30.4	26
17*	-.313	31	31.3	21
40*	-.399	32	30.8	23
7*	-.511	33	29.8	28

*Teachers who were consistent over time and who were used in the top nine, bottom nine analysis.

**Teachers who taught the class as group.

Table 12

Top Nine and Bottom Nine Teachers Ranked on the Basis of Pre-Scores

High Group			Low Group		
Teacher #	Pre-Mean	Rank	Teacher #	Pre-Mean	Rank
24	5.15	3	7	3.49	18
16	4.31	12	40	3.57	17
33	5.25	2	17	4.78	5
27	4.27	13	21	3.79	16
3	4.65	7	31	3.81	15
19	4.57	9	8	4.57	9
22	4.13	14	1	4.80	4
18	4.57	9	20	4.39	11
39	5.28	1	9	4.76	6

Table 13

Absolute Analysis: Mean residual scores from a 3 (teacher competence) x 3 student aptitude analysis of variance done separately by grade level

Grade 3

Teacher Competence	Student Aptitude		
	High	Middle	Low
High	.30	.02	.03
Middle	.09	.04	-.14
Low	-.19	-.20	-.27

Grade 4

Teacher Competence	Student Aptitude		
	High	Middle	Low
High	.35	.21	.14
Middle	.80	.00	-.15
Low	-.01	-.27	-.29

Table 14

Relative Analysis: Mean residual gain scores from two 3 (teacher competence) x 3 (student aptitude analyses of variance done separately by grade level

Grade 3

Teacher Competence	Student Aptitude		
	High	Middle	Low
High	.34	.18	-.02
Middle	.12	-.001	-.13
Low	-.20	-.18	-.29

Grade 4

Teacher Competence	Student Aptitude		
	High	Middle	Low
High	.37	.20	.14
Middle	.07	-.01	-.12
Low	-.11	-.28	-.29

Table 15

Percent of Students Who Exceed Expected Performance
In Each Room by Sex and Achievement Level

	TOTAL		MALES		FEMALES		HI ACHV.		MIDDLE ACHV.		LOW ACHV.	
TEACH.	N	%ABV	N	%ABV	N	%ABV	N	%ABV	N	%ABV	N	%ABV
01	22	36	11	36	11	36	7	38	7	14	8	50
02	18	50	11	64	7	28	6	33	5	60	7	57
03	24	58	13	62	11	54	9	55	7	57	8	62
04	26	50	15	60	11	36	8	62	8	38	10	50
05	20	28	10	30	10	30	7	28	7	28	6	33
07	14	14	7	0	7	28	5	0	4	0	5	40
08	20	40	9	44	11	27	7	57	7	28	6	33
09	13	31	7	14	6	50	4	25	4	0	5	60
10	22	59	10	60	12	58	7	43	8	62	7	71
11	23	42	10	50	13	38	8	50	8	38	7	43
12	22	50	12	33	10	70	8	62	7	43	7	43
13	22	59	12	66	10	50	7	57	8	38	7	86
14	26	42	12	50	14	43	9	33	9	33	8	75
16	25	76	11	64	14	86	7	100	10	70	8	50
17	22	21	13	8	9	44	7	14	8	25	7	29
18	25	64	14	64	11	63	7	71	9	77	9	44
19	24	62	14	64	10	60	7	71	10	50	7	71
20	21	33	10	30	11	36	7	28	7	43	7	28
21	20	30	12	41	8	12	7	0	8	38	5	60
22	24	66	13	62	11	73	8	62	10	60	6	83
23	12	42	8	38	4	50	5	0	5	60	2	100
24	33	70	12	83	21	62	11	54	12	75	10	80
26	19	58	11	63	8	50	5	60	8	50	6	66
27	20	50	10	60	10	40	7	57	7	57	6	33
29	22	46	7	28	15	53	7	57	7	43	8	37
30	22	64	8	62	14	64	7	28	8	75	7	85
31	19	42	12	50	7	28	6	50	7	28	6	50
33	17	76	4	50	13	84	6	83	8	87	3	33
36	19	37	14	50	5	20	6	0	7	57	6	50
38	21	43	9	44	12	42	7	43	7	28	7	57
39	29	59	18	55	11	63	9	55	10	70	9	55
40	17	30	9	22	8	38	4	0	7	28	6	50
41	26	50	11	73	15	33	6	50	10	20	10	80

Table 16

List of Variables Included in Discriminant Analysis

Number of students
Time in whole class instruction
Time spent going over homework
Number of direct questions
Number of process questions
Number of correct responses
Number of wrong responses
Number of don't know responses
Number of no response
Teacher praise after student response
Teacher negation following incorrect student response
Teacher repeats question
Created work process feedback
Created work feedback
Afforded work: unknown feedback
Teacher gives behavioral warning
Teacher afforded work contact
Teacher afforded behavioral contact
Total teacher afforded contact
Student created work
Total student created
Total teacher process feedback

Table 17

Summary Table from the First Discriminant Analysis

Variable	F Value to Enter or Remove
Number of students	37.24
Wrong answers	9.37
Time teaching whole class	5.13
Total student initiated work	4.27
Teacher afforded work contact	4.55
Teacher behavior warning	3.19
Number of process questions	2.92
Teacher negation following incorrect student response	2.03

Table 18

Summary Table from the Second Discriminant Analysis with the Varable Number of Students Removed

Variable	F to Enter or Remove
Total student created contact	15.75
Teacher warning	7.53
Teacher praise	5.11
Process questions	5.00
Teacher negation following incorrect answer	2.31
Time spent teaching whole class	2.47
Number of correct answers	2.82
Number of no response	3.25
Time spent going over homework	3.45
Teacher afforded work contact: feedback unknown	3.51

Table 19

Summary Table from the Third Discriminant Analysis with the Variables Number of Students in Class and Time Spent Teaching the Class as a Whole Removed

Variable	F to Enter or Remove
Total student created	15.75
Teacher warning	7.53
Teacher praise	5.11
Process questions	4.99
Negation following incorrect response	2.31
Correct answers	2.22
No response	2.54
Teacher afforded work: feedback unknown	2.61
Time going over homework	2.78

Table 20

Prediction Results from the Three Discriminant Analyses

First Analysis

Actual Group	No. of Cases*	Predicted High	Predicted Low
High	57	44 (77.2%)	13 (22.8%)
Low	58	9 (15%)	49 (84.5%)

Percent of Cases Correctly Classified: 80.87% **

Second Analysis

Actual Group	No. of Cases	Predicted High	Predicted Low
High	57	48 (84.2%)	9 (15.89%)
Low	58	16 (27.6%)	42 (72.4%)

Percent of Cases Correctly Classified: 78.26%

Third Analysis

Actual Group	No. of Cases	Predicted High	Predicted Low
High	57	49 (86%)	8 (14%)
Low	58	17 (29.3%)	41 (70.7%)

Percent of Cases Correctly Classified: 78.26%

*There were 57 observations made on the top 9 teachers and 58 completed in the bottom 9.
**When means rather than individual observations are used, the dissemination is perfect.

References

Berliner, D. Impediments to the study of teacher effectiveness. An expanded version of a paper presented at the meetings of the National Association for Research in Science Teaching, Los Angeles, California, March 18, 1975.

Brophy, J. Stability in teacher effectiveness. A paper read at the annual meeting of the American Educational Research Association. Chicago, Illinois, 1972.

Brophy, J. and Evertson, C. Process-Product correlations in the Texas teacher effectiveness study: Final report (Res. Rep. 74-4). Austin, Texas: Research and Development Center for Teacher Education, 1974. (ERIC No. ED 091 394).

Brophy, J. and Good, T. Brophy-Good System (Teacher-Child Dyadic Interaction). In A. Simon and E. Boyer, Mirrors for Behavior: An Anthology of Observation Instruments Continued, 1970 supplement. Volume A. Philadelphia: Research for Better Schools, Inc., 1970.

Dunkin, M. and Biddle, B. The Study of Teaching. New York: Holt, Rinehart and Winston, 1974.

Emmer, E. Classroom Observation Scales. Austin, Texas: Research and Development Center for Teacher Education, 1973.

Good, T. and Grouws, D. Teacher rapport: some stability data. Journal of Educational Psychology, 1975, 67, 179-182.

Good, T. and Power C. Designing successful classroom environments for different types of students. Journal of Curriculum Studies (in press).

Grouws, D. and Good, T. The status of manipulative materials in elementary school mathematics instruction. Technical Report No. 107. Center for Research in Social Behavior, University of Missouri--Columbia, Columbia, Mo., 1975.

Kleinfeld, J. Instructional style and the intellectual performance of Indian and Eskimo students. Final Report, Project No. 1-J-027, Contract No. OE C-X-71-0019 (057). U. S. Department of Health, Education and Welfare. Office of Education, 1972.

Kounin, J. Discipline and Group Management in Classrooms. New York: Holt, Rinehart and Winston, 1970.

Mitzel, H. and Gross, C. The development of pupil growth criteria in studies of teacher effectiveness. Educational Research Bulletin, 1958, 37, 178-187.

Morsh, J. L. and Wilder, E. W. Identifying the effective instructor: a review of quantitative studies, 1900-1952. Research Bulletin No. AFPTRC-TR-54-44. San Antonio, Texas: ASAF Personnel and Training Center.

Rabinowitz, W. and Rosenbaum, I. Failure in the prediction of pupil-teacher rapport. Journal of Educational Psychology, 1958, 49 93-98.

Rosenshine, B. and Furst, N. Current and future research on teacher performance criteria. In Smith, B. (ed.), Research on Teacher Education: A Symposium. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971.

Soar, R. S. An integrative approach to classroom learning. Public Health Service Grant No. 5-R11 MH 01096 and National Institute of Mental Health Grant No. 7-R11 MH 02045. Temple University, Philadelphia, Pennsylvania, 1966.

Stephens, J. M. The Process of Schooling. New York: Holt, 1967.

Torrance, E. and Parent, E. Characteristics of Mathematics Teachers that Affect Students' Learning. Cooperative Research Project No. 1020. Minnesota School Mathematics and Science Center, Institute of Technology, University of Minnesota, 1966.

Appendix A

Coding Instruments Used to Collect Process Data

Overview

Figure 1 presents the basic coding instrument: The Brophy-Good Dyadic Observation Instrument. Definitions of coding variables in this system follow Figure 1.

Figure 2 presents the time categories and definitions that were used to describe the way in which mathematics instructional time was used.

Figure 3 presents high-inference variables that were coded at the end of each observation. Variable definitions follow Figure 3.

Figure 4 presents high inference variables that were coded when the teacher was presenting material to the class or to a group. Variable definitions follow Figure 4.

Figure 5 presents variables that were coded when the teacher was supervising seat work or board work. Definitions follow Figure 5.

Figures 6 and 7 present instruments that were used whenever the teacher used concrete instructional material or assigned homework.

Figure 1

1. Child ______ 2. Date ______ 3. Start ______ 4. Stop ______
6. Activity ______ 7. Attendance ______ 8. Observer ______

TIME | CHILD | QUESTION | ANSWER | RESPONSE OPPORTUNITIES: TERMINAL FEEDBACK | SUSTAINING | WORK | PROCED. | WORK | SEA.

EXPECTATION:

REMARKS:

Group Questions

Board Work

Brief Definitions of Variables Coded in Brophy-Good Dyadic Interaction System

The coding sheet (see Figure 1) used the definitions that appear immediately below. Definitions are presented in the order that they appear on the coding sheet. For an extended discussion of these definitions and coding examples see Brophy and Good (1970).

Student Initiated Question: A student asks the teacher a question in a public setting.

Reading or Recitation: Student is called upon to read aloud, go through an arithmetic table, etc.

Discipline Question: The discipline question is a unique type of direct question in which the teacher uses the question as a control technique, calling on the child to force him to pay better attention rather than merely to provide a response opportunity in the usual sense.

Direct Question: Teacher calls on a child who is not seeking a response opportunity.

Open Question: The teacher creates the response opportunity by asking a public question, and also indicates who is to respond by calling on an individual child, but he chooses one of the children who has indicated a desire to respond by raising his hand.

Call Outs: Response opportunities created by children who call out answers to teachers' questions without waiting for permission to respond.

Process Question: Requires students to explain something in a way that requires them to integrate facts or to show knowledge of their interrelationships. It most frequently is a "why?" or "how?" question.

Product Question: Product questions seek to elicit a single correct answer which can be expressed in a single word or short phrase. Product questions usually begin with "who?," "what?," "when?," "where?" "how much? " or "how many?"

Choice Question: In the choice question the child does not have to produce a substantive response but may instead simply choose one of two or more implied or expressed alternatives.

Self-Reference Question: Asks the child to make some non-academic contribution to classroom discussion ("show and tell," questions about personal experiences, preferences, or feelings, requests for opinions or predictions, etc.).

Opinion Questions: Much like self-reference (i.e. no one correct answer) except that they seek a student opinion on an academic topic (is it worth putting a man on the moon?).

Correct Answers: If the child answers the teacher's question in a way that satisfies the teacher, the answer is coded as correct.

Part-Correct Answers: Part-correct answers are answers which are correct but incomplete as far as they go or answers which are correct from one point of view but not the answer that the teacher is looking for.

Incorrect Answers: Responses coded as incorrect answers are those in which the child's response is treated as simply wrong by the teacher.

Don't Know: Student verbally says "I don't know" (or its equivalent) or nonverbally indicates that he doesn't know (shakes head).

No Response: Student makes no response (verbally or non-verbally) to teacher question.

Praise: Praise refers to the teacher's evaluative reactions which go beyond the level of simple affirmation or positive feedback by verbally complimenting the child.

Affirmation of Correct Answers: Affirmation is coded when the teacher indicates that the child's response is correct or acceptable.

Summary: Teacher summarizes the student answer (generally as part of the affirmation process).

No Feedback Reaction: If the teacher makes no verbal or nonverbal response whatever following the child's answer to the question, he is coded for no feedback reaction.

Negation of Incorrect Answers: Simple provision of impersonal feedback regarding the incorrectness of the response, and not going further than this by communicating a personal reaction to the child. As with affirmation, negation can be communicated both verbally ("No," "That's not right," "Hmm-mm") and non-verbally (shaking the head horizontally).

Criticism: Evaluative reactions that go beyond the level of simple negation by expressing anger or personal criticism of the child in addition to indicating the incorrectness of his response.

Process Feedback: Coded when the teacher goes beyond merely providing the right answer and discusses the cognitive or behavioral processes that are to be gone through in arriving at the answer.

Gives Answer: This category is used when the teacher gives the child the answer to the question, but does not elaborate sufficiently to be coded for process feedback.

Asks Other: Whenever the child does not answer a teacher question and the teacher moves to another child in order to get the answer to that same question, the teacher's feedback reaction is coded for asks other.

Call Out: The call out category is used when another child calls out the answer to the question before the teacher has a chance to act on his own.

Repeats Question: Teacher asks a question, waits some time without getting the correct answer, and then repeats the question to the same child.

Rephrase or Clue: In this feedback reaction, the teacher sustains the response opportunity by rephrasing the question or giving the child a clue as to how to respond to it.

New Question: The teacher asks a new question when she requires an answer that is different from the original question, although it may be closely related. A question requiring a new answer is coded as a new question.

Expansion: Teacher statements that ask the student to provide more information (I think I understand but give me . . .).

Dyadic Teacher-Child Contacts

The preceding material has dealt primarily with the coding of response opportunities and reading and recitation turns. Dyadic teacher-child contacts differ from response opportunities and reading and recitation turns in that the teacher is dealing privately with one child about matters idiosyncratic to him rather than publicly about material meant for the group or class as a whole.

Dyadic teacher-child contacts are divided into procedural contacts, work related contacts, and behavioral or disciplinary contacts. They are also separately coded according to whether they are initiated by the teacher (teacher-afforded) or by the child (child created). The coding also reflects certain aspects of the teacher's behavior in such contacts.

Work-Related Contacts

Work-related contacts include those teacher-child contacts which have to do with the child's completion of seat work or homework assignments. They include clarification of the directions, soliciting or giving help concerning how to do the work, or soliciting or giving feedback about work already done. Work-related interactions are considered child-created if the child takes it upon himself to bring his work up to the teacher to talk to him about it or raises his hand or otherwise indicates that he wants to discuss it with him. Work-related interactions are coded as teacher-afforded if the teacher gives feedback about work when the child has not solicited it (the teacher either calls the child to come up to his desk or goes around the room making individual comments to the students. Created contacts are not planned by the teacher and occur solely because the child has sought him out; afforded contacts are not planned by the child and occur solely because the teacher initiates them. Separate space is provided for coding created and afforded work related interactions on the coding sheets, and the coder indicates the nature of an individual dyadic contact by where he codes the interaction.

In addition to noting the interaction as a work interaction and as an interaction which is child-created or teacher-afforded, the coder also indicates the nature of the teacher's feedback to the child during the interaction. He indicates this by using one or more of the five columns provided for coding teacher's feedback in work related interaction: praise (++), process feedback (pcss), product feedback (fb), criticism (--), or "don't know" (?). The first four of these categories have the same meaning as they have in other coding of teacher feedback. The additional "don't know" category is added for this coding because frequently the individual teacher-child interaction that occurs in the dyadic contacts will be carried on in hushed tones or across the room from the coder where he cannot hear the content of the interaction. In such cases, the coder notes the occurrance of the work related interaction and the fact that it was either teacher afforded or child created, but he enters the number in the "don't know" column.

Procedural Contacts

The category of procedural contacts includes all dyadic teacher-child interaction which is not coded as work-related contacts or as behavioral contacts. Thus it includes a wide range of types of contacts, most of which are initiated on the basis of the immediate needs of the teacher or child involved. Procedural contacts are created by the child for such purposes as seeking permission to do something, requesting needed supplies or equipment, reporting some information to the teacher (tattling on other children, calling his attention to a broken desk or pencil, etc.).

Three categories for coding teacher's response are provided: praise (++), feedback (fb), and criticism (--). Praise and criticism have the same meaning here as elsewhere.

Behavioral Contacts

Behavioral contacts are coded whenever the teacher makes some comment upon the child's classroom behavior. They are subdivided into praise, warnings, and criticism. Praise and criticism are coded as described above.

Warning

Usually teachers' warnings will occur in situations in which the child is doing something that is not necessarily or always prohibited but which is troublesome at the moment. In such instances the teacher will single out the child to inform him that his present behavior is inappropriate, but will do so without communication of rejection or anger as in criticism.

Figure 2*

Instructional Time Variables and Definitions

Transition: The time between teachers' initial call for mathematics instruction to begin and when instruction actually commenced.

Total class time: Time between initial instruction and end of math period.

Time teacher taught class as a whole: Time teacher spent working with the entire class.

Time teacher taught group: Time teacher spent working with groups of students. .

Time teacher taught individuals: Time teacher spent working with individual students.

Time going over homework: Amount of class time that was used in reviewing previously assigned homework.

Review time: Amount of class time used to review previously presented facts and concepts that were not part of an explicit homework assignment.

Development: Amount of time devoted to increasing comprehension of skills, concepts.

Drill: Time spent in practice where rapid recall and/or accurate processing is given special attention. Time on practice problems that students work at their desk and that will be checked later in the period are also subsumed in this category.

Homework Practice (Drill) and Homework Development: Homework is coded after the teacher makes an explicit assignment of work that will not be reviewed until the following day. Drill and development are coded as described above.

*Coder agreement in coding these categories was virtually perfect. Time codes were entered in the left hand column in the Brophy-Good Instrument that was presented in Figure 1.

Figure 3

Summary Variables Coded at the End of Each Observation

Organization	___1	___2	___3	
Alerting	___1	___2	___3	
Accountability	___1	___2	___3	
Climate	___1	___2	___3	___4
Managerial effectiveness	___1	___2	___3	___4

A single high inference code was entered for each of the five variables at the end of each observation.

The scale points were defined as follows:

Organization
1) Well structured--students knew what to do.
2) Some students understood, but not all.
3) Much confusion and wasted time.

Alerting
1) Used a variety of group alerting cues, no negative cues. Almost all the students were involved and alert.
2) Used several group alerting cues, 75-25% of students were involved.
3) Used few cues with <25% student involvement.

Accountability
1) Checked 50-100% of students
2) 25-50%
3) <25%

Managerial Effectiveness
1) Teacher well organized, class flowed smoothly, one or two minor disruptions, little wasted time.
2) Three or four minor disruptions, no serious disruptions, teacher in control of class movement.
3) Four or five minor disruptions, or serious disruptions, Teacher generally in control.
4) Five or more minor disruptions and/or one or more serious disruptions, flow uneven, teacher unable to direct class.

Climate
1) Positive, interesting, relaxed, unpressured.
2) More interesting than work.
3) Neutral
4) Work is work (grind).

Coder agreement was generally good on these scales (percent of time within one scale point). The following percent of agreement codes were obtained: organization .77; alerting .69; accountability .85; managerial effectiveness .75; and climate 1.0. Coder agreement on the first three variables was based upon thirteen independent observations during pilot work. The scales managerial effectiveness and climate were added toward the end of pilot work and were based upon but four comparisons.

General concepts for coding accountability and alerting were derived from Kounin's (1970) work. He provides the following general descriptions of concepts.

Alerting

Group alerting refers to the degree to which a teacher attempts to involve nonreciting children in the recitation task, maintain their attention, and keep them "on their toes" or alerted. Anything the teacher does that indicates an overt effort on her part to get more than the reciter attentive and involved was considered a group alerting cue.

Positive group alerting cues are those behaviors of a teacher that keep nonreciters on their toes while another child is reciting or before the selection of a new reciter. Positive group alerting cues were:

1. Any method used to create "suspense" before calling on a child to recite: pausing and looking around to "bring children in" before selecting a reciter; saying, "Let's see now, who can . . ." before calling on a reciter.

2. Keeping children in suspense in regard to who will be called on next; picking reciters "randomly" so that no child knows whether he will be called on next or not.

3. Teacher calls on different children frequently or maintains group focus: intersperses "mass unison" responses; says, "Let's put our thinking caps on; this might fool you;" asks group for show of hands before selecting a reciter.

4. Teacher alerts nonperformers that they might be called on in connection with what a reciter is doing: They may be called on if reciter makes a mistake; presignals children that they will be asked about recitation content in the immediate future.

5. Teacher presents new, novel, or alluring material into a recitation (a high attention value prop or issue).

Negative group alerting cues were those behaviors of a teacher during a child's recitation, or preceding the selection of a new reciter, that reduced the involvement of nonreciters in a recitation session. These produced a lower degree of alertness on the part of nonreciters than that obtaining in ordinary, routine recitation sessions.

Negative group alerting cues were:

1. The teacher changes the focus of her attention away from the group and becomes completely immersed in the performance of the reciter; or directs a new question and subsequent attention to a single new reciter only, without any overt sign of awareness that there is a group.

2. The teacher prepicks a reciter or performer before the question is even stated.

3. The teacher has reciters perform in a predetermined sequence of turns. That is, children know beforehand that they are to read from left to right with the child at the far left reading first, then the child next to him, then the child next to him or her, and so on. (In contrast to a random selection of reciters, a child in this sequence knows ahead of time when he is not and when he is going to be called upon to recite.)

<u>Accountability</u>

Accountability refers to the degree to which the teacher holds the children accountable and responsible for their task performances during recitation sessions. This entails her doing something to get to know what the children are actually doing and to communicate to the children in some observable manner that she knows what they are doing. The degree to which she goes out to obtain this knowledge and to communicate it, is the degree to which she holds the children in the group accountable for their performances. <u>A teacher's accountability score is the number of children she makes accountable</u>.

1. Teacher requires children to recite in unison while the teacher shows signs of actively attending to the recitation.

2. Teacher brings other children into the performance of a child reciting. (Teacher says, "Jimmy, you watch Johnny do that problem and then tell me what he did right or wrong.")

3. Teacher asks for the raised hands of children who are prepared to demonstrate a performance and requires some of them to demonstrate.

4. Teacher circulates and checks products of nonreciters during a child performance.

5. Teacher requires a child to demonstrate and checks his performance.

The organizational scale was simply a high inference rating to the structure of learning assignments that took place. Did students understand their options? The managerial effectiveness scale allowed coders to react to the number of disturbances that had occurred during the math period. The climate variable focused upon the milieu in which work was completed. In general this code attempted to measure the degree to which the work was completed in a pleasant atmosphere. All three of these variables were high inference variables used with only the scale anchor points.

Accountability and Alerting as used here were also high inference variables since they represented the coders general estimate for the period as a whole. Another scale (to be described later) contained more focused observation with the accountability and alerting scale. Four of the five high inference variables described here did not discriminate between high and low classrooms.

However accountability and alerting codes in the more focused observation (to be described later) did prove to be useful discriminators. Part of the problem in the summary usage may have been that the three point scale was too limited to describe an entire observation period and modality confusion (for example the alerting scale forces the coder to discriminate both teacher behavior (number and type of alerting cues) as well as effect upon student behavior (how attentive)).

Figure 4

Date__________ I.D.__________

Observation________________

High Inference Codings

Time Code	Homework			Attention					Clarity					Enthusiasm					Presentation				
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

Figure 4 shows the high inference codes that were made at _two minute intervals_ during portions of the mathematics lesson when the teacher was working with the class as a whole in recitation, development, review, and drill (if it involved frequent teacher-pupil agreement _public_ discussion). Coder agreement based on 30 ratings (five ratings on six different lessons) was acceptable on all scales: Homework 83; Attention 73; Clarity .80; Enthusiasm .80; and Presentation .70.

Homework ratings indicated the extent to which the time spent in introducing homework or reviewing homework focused upon development (comprehension, process) vs. product (right answer, accurate processing). _Lower scores on this scale indicated more process orientation._

The attention, clarity, enthusiasm, and presentation were all taken from Emmer (1973). Definitions of these variables follow:

Attention

Attention as defined for this scale refers to pupil orientation toward the teacher, the task at hand, or whatever classroom activities are appropriate. If a pupil is attending to inappropriate activities, or is engaged in self-directed behavior when he is supposed to be engaged in a class activity, his behavior is not considered attentive. Therefore, you should look for behavior that is focused upon or engaged in whatever activity is appropriate. At times, it will be difficult to determine whether a student is being attentive, such as when the teacher presents information and the student sits facing the teacher, with no observable behaviors indicating inattention. In such an instance, the pupil is considered attentive until he provides a behavioral indicator to the contrary.

Observers enter one of the following codes every two minutes during times when the teacher was presenting material to the class.

1. Fewer than half of the students are attentive most of the time.

2. One-half to three-fourths of the students appear attentive most of the time; the remainder are attentive only some of the time.

3. Most of the students are attentive, but several (four to six) are attentive only some of the time.

4. Nearly all students are attentive, but a few (one, two or three) are attentive only some of the time.

5. All the students are attentive most of the time.

Note--the phrase "most of the time" means at least 75% of the time the observer checks the pupils for attentiveness.

Clarity

Clarity refers to the degree to which the teacher's presentation of material and his substantive interactions with students are understood by them. Low clarity means that the teacher is "over their heads" and is confusing to the pupils.

1. Very low clarity. Pupils seem very confused by the presentation. The teacher cannot answer the pupils' questions, or answers them in an unclear manner by using concepts and terms the pupils are apparently unfamiliar with or by being overly complex and ambiguous.

2. Low clarity. Between very low and moderate.

3. Moderate clarity. The teacher seems to be understood by most pupils, but not all of the time. Sometimes the teacher is confusing and vague.

4. High clarity. Between moderate and very high.

5. Very high clarity. The teacher's explanations are easy to understand and pupil questions are adequately answered. The teacher seems aware of the pupils' levels, sensing problems they are having or may have.

Enthusiasm

This scale is used to judge the extent to which the teacher displays interest, vitality, and involvement in his subject and his instruction.

1. Very low enthusiasm. The teacher's behavior is lethargic, dull, routine; a minimum of vocal inflection, gesturing, movement, or change in facial features. The teacher appears to lack interest in what he is doing.

2. Low enthusiasm. Between very low and moderate.

3. Moderate enthusiasm. Occasionally the teacher seems interested and involved; some display of activity, such as gesturing. Sometimes the teacher is dull, routine, and lacking in vigor.

4. High enthusiasm. Between moderate and very high.

5. Very high enthusiasm. The teacher is stimulating, energetic, and very alert. He seems interested and involved in what he is teaching; moves around, gestures, inflects voice.

Presentation

This scale measures only one type of behavior. The observer's task is to estimate the relative amount of class time occupied by teacher presentation of substantive information. By teacher presentation is meant substantive (content oriented) verbal or non-verbal behavior that provides information, and does not imply or require pupil response, nor evaluate pupil behavior. Thus, teacher questions, procedural directsion, praise, and criticism are not instances of teacher presentation. Lecture, reading to the class, answering pupil questions, and any other activity in which the teacher gives information will be an instance of teacher presentation.

1. Teacher presentation occurs 0-20% of the period.

2. Teacher presentation occurs 20-40% of the period

3. Teacher presentation occurs 40-60% of the period.

4. Teacher presentation occurs 60-80% of the period.

5. Teacher presentation occurs 80-100% of the period.

Figure 5

Date__________ I.D.__________

Observation____________________

Seatwork - Boardwork

Time	Activity Code	Accountability Code	Alerting Code	Involvement Code
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3
	__________	_1 _2 _3 _4	_1 _2 _3 _4	_1 _2 _3

Figure 5 shows variables that were coded during the times when the teacher was supervising seatwork and boardwork activities. To reiterate, the Brophy-Good Dyadic System was used at all times. Supplementary rating information was collected every two minutes with either the scale shown in Figure 4 or Figure 5 depending upon the nature of the activity.

The accountability code and alerting codes are based upon Kounin's work and the general concepts have been discussed above. On this scale; however, a rating is being made every several minutes rather than a single global estimate (as on the summary sheet).

Reliability for accountability and alerting were adequate (.77 and .73). General scale points were defined in the following ways for accountability and alerting.

Accountability

1 = from ½ to every student held accountable.

2 = from 25-50% held accountable.

3 = less than 25%.

4 = no accountability efforts.

Alerting

1 = 3 or more alerting behaviors during time interval.

2 = 2 or more alerting behaviors.

3 = 1 alerting behavior.

4 = no alerting behavior.

The student involvement code was also adapted from Kounin (1970). The scale definitions for this variable follows.

1. Definitely in the assigned work. To be coded as definitely working, a student had to exhibit overt signs that he was "in" the prescribed activity (writing in the prescribed workbook; performing or volunteering to perform in recitation; in a clear posture of attending or reading, etc.).

2. Probably in the assigned work. For the behavior to be coded in this category, the child had to be in a posture from which it could reasonably be inferred that he was "in" the work (looking at or having proper props before him, oriented physically as though he could be thinking or listening) but not actually writing or otherwise showing clear signs of being "in."

3. Definitely out of the assigned work. This was coded when the child showed no actual or postural signs of being "in" or showed clear signs that he was "out" of the prescribed activity--was attending to, or doing something other than the designated task.

In using this scale, coders scaned the classroom every five minutes and rated the number of students that fell into each category.

Figure 6

Date ________

ID. # ________

Observation # ________

Materials

Name or Description ________________________

Topic being studied ________________________

Material type. Software ___ Hardware ___

Production: Teacher-made ___ Commercially-made ___

Usage: Teacher only ___ Student only ___ Both ___ est. ratio.

Context: Developmental ___ Practice ___ Other (describe)

Competitive. Heavy ___ Some ___ None ___

Appropriate: Yes ___ No ___

Judgement concerning instructional usage

1	2	3	4	5
very ineffective	ineffective	average	effective	very effective

Judgement concerning affective response

1	2	3	4	5
very ineffective	ineffective	average	effective	very effective

Figure 6 presents the form that was used for recording information about the type of material that was used in instructional activity (if any) and how it was developed. Material was classified according to type, production, usage, and instructional context. Judgments about the competitiveness of the situation as well as the extent to which the material was related to the mathematics topic (e.g. mathematics games that follow instruction may or may not be directly related to the instructional unit). Judgments about teachers' effectiveness in using material and students' affective responses were also made.

Material type was classified as either hardware or software. Hardware was defined as physical equipment used to supplement the textbook (balance beams, commercial games, abaci, blocks, etc.). Software was defined as printed material other than the textbook (task cards, flash cards, flannel boards, etc.). Distinctions were also made with regard to whether the material was teacher made or commercially produced and the extent to which the teacher, students, or both used the material.

The developmental portion of a mathematics period is that part of a lesson devoted to increasing comprehension of skills, concepts, and other facets of the mathematics curriculum. For example, in the area of skill development, instruction focusing on why an algorithm works, how certain skills are interrelated, what properties are characteristic of a given skill, and means of estimating correct answers, should be considered part of developmental work. In the area of concept development, developmental activities would include initial instruction designed to help children distinguish the given concept from other concepts. Also included would be the associating of a label with a given concept.

Attempts to extend ideas and facilitate transfer of ideas are a part of developmental work.

The practice portion of a mathematics period is that part of a lesson where rapid recall and accurate processing are given special attention and importance. Generally comprehension is assumed or overlooked at this point. The activities often center on polishing previously encountered ideas.

A competitive situation is defined to be any situation where the intent or outcome is for a student or a group of students to explicitly exhibit greater proficiency at a given task or set of tasks than another student or group of students. The main intent is not to get teacher "attention." This rules out: (1) competition for teacher attention, (2) normal question-answer type session.

The appropriateness of a set of materials must be judged in an instructional context. That is, a set of materials is judged to be appropriate for use in a mathematics classroom if and only if the material embodies in a meaningful way the given idea being studied. Use of a game or device which the children find enjoyable and which involves some mathematics is not appropriate unless the mathematics used corresponds to the mathematics to be taught at that given point in time.

This scale was developed immediately prior to the collection of observation data and very little reliability data exist. However what data do exist (3 independent observations of teachers using concrete materials) are supportive. Excellent agreement was reached on the distinctions from material type through topic appropriateness. Ratings on the two judgment scales (appropriate usage and affective response) were within 2 scale points.

Figure 7

Date ____________

ID. # ____________

Observation # ____________

Home Work

1. Is homework assigned? Yes ___ No ___

2. Same assignment to all students? Yes ___ No ___

3. Time provided in class to work on homework? Yes ___ No ___

4. If so, how much time? ______________________________

5. Describe homework assignment:

	Yes	No
A. textbook	___	___
B. workbook	___	___
C. commercial ditto	___	___
D. teacher made ditto	___	___

Figure 7 presents the information that was collected when homework assignments were made. Coders obtained near perfect agreement using this scale.

Appendix B: Climate Scale

Name ____________________ Sex __________

Grade __________ Teacher __________ School __________

1. Do you like to be in this class?

   Always ____ Most of the time ____ Sometimes ____ Never ____

2. Do you have much fun in this class?

   Always ____ Most of the time ____ Sometimes ____ Never ____

3. Do most of your close friends like the teacher?

   Always ____ Most of the time ____ Sometimes ____ Never ____

4. Does the teacher help you enough?

   Always ____ Most of the time ____ Sometimes ____ Never ____

5. Do you learn a lot in this class?

   Always ____ Most of the time ____ Sometimes ____ Never ____

6. Do you ever feel like staying away from this class?

   Always ____ Most of the time ____ Sometimes ____ Never ____

7. Are you proud to be in this class?

   Always ____ Most of the time ____ Sometimes ____ Never ____

8. Do you always do your best in this class?

   Always ____ Most of the time ____ Sometimes ____ Never ____

9. Do you talk in class discussions in this class?

   Always ____ Most of the time ____ Sometimes ____ Never ____

10. Are most of the students in this class friendly to you?

    Always ____ Most of the time ____ Sometimes ____ Never ____