Occasional Paper No. 73

TEACHER BEHAVIOR AND STUDENT ACHIEVEMENT

Jere Brophy
and
Thomas L. Good

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

April 1984

This work is sponsored in part by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Program for Teaching and Instruction of the National Institute of Education, United States Department of Education. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-81-0014).

Institute for Research on Teaching

The **Institute for Research on Teaching** was founded at Michigan State University in 1976 by the National Institute of Education. Following a nationwide competition in 1981, the NIE awarded a second contract to the IRT, extending work through 1984. Funding is also received from other agencies and foundations for individual research projects.

The IRT conducts major research projects aimed at improving classroom teaching, including studies of classroom management strategies, student socialization, the diagnosis and remediation of reading difficulties, and teacher education. IRT researchers are also examining the teaching of specific school subjects such as reading, writing, general mathematics, and science, and are seeking to understand how factors outside the classroom affect teacher decision making.

Researchers from such diverse disciplines as educational psychology, anthropology, sociology, and philosophy cooperate in conducting IRT research. They join forces with public school teachers, who work at the IRT as half-time collaborators in research, helping to design and plan studies, collect data, analyze and interpret results, and disseminate findings.

The IRT publishes research reports, occasional papers, conference proceedings, a newsletter for practitioners, and lists and catalogs of IRT publications. For more information, to receive a list or catalog, and/or to be placed on the IRT mailing list to receive the newsletter, please write to the IRT Editor, Institute for Research on Teaching, 252 Erickson Hall, Michigan State University, East Lansing, Michigan 48824-1034.

Abstract

This paper, prepared as a chapter for the *Handbook of Research on Teaching* (third edition), reviews correlational and experimental research linking teacher behavior to student achievement. It focuses on research done in K-12 classrooms in 1973-1983, highlighting several large-scale, programmatic efforts. Attention is drawn to design, sampling, measurement, and context (grade level, subject matter, student socioeconomic status) factors that must be taken into account in interpreting this research and in comparing the findings of different studies. Topics covered include opportunity to learn/content covered, teacher expectations/role definitions/time allocations, classroom management/student engaged time, success level/academic learning time, active group instruction by the teacher, group size, presentation of information (structuring, sequencing, clarity, enthusiasm), asking questions (difficulty level, cognitive level, wait-time), selecting respondents, providing feedback, and handling seatwork and homework assignments.

Contents

Contents (continued)

Contents (continued)

TEACHER BEHAVIOR AND STUDENT ACHIEVEMENT[1]

Jere Brophy and Thomas L. Good[2]

This paper reviews process-product (also called process-outcome) research linking teacher behavior to student achievement. Within this, the paper stresses (1) teacher behavior over other classroom process variables (students' interactions with peers, curriculum materials, computers, etc.) and (2) student achievement gain over other product variables (e.g., personal, social, or moral development).

The research to be discussed concerns teachers' effects on students, but it is a misnomer to refer to it as "teacher effectiveness" research, because this equates "effectiveness" with success in producing achievement gain. What constitutes "teacher effectiveness" depends on definition, and most definitions include success in socializing students and promoting their affective and personal development in addition to success in fostering their mastery of formal curricula. Consequently, we have avoided the term "teacher

---

[1]This paper appears as a chapter in the *Handbook of Research on Teaching* edited by M.C. Wittrock and to be published by MacMillan, New York, NY (in press). In addition to assigned reviewers David Berliner and Virginia Koehler, the authors wish to thank Linda Anderson, Christopher Clark, Mary Rohrkemper and (especially) Barak Rosenshine for their comments on earlier drafts, and June Smith for her assistance in manuscript preparation.

[2]Jere Brophy is co-director of the IRT and a professor in MSU's Department of Teacher Education. Thomas L. Good is research associate at the Center for the Study of Social Behavior and a professor in the Department of Curriculum and Instruction at the University of Missouri-Columbia.

effectiveness" in titling this paper and describing the research, although we use the more neutral term "teacher effects."

Developments in this field have been well documented in previous handbook chapters (Medley & Mitzel, 1963; Rosenshine & Furst, 1973), and in volumes by Rosenshine (1971) and by Dunkin and Biddle (1974). This paper, therefore, builds on these earlier reviews without overlapping them unnecessarily. It attempts to be comprehensive in covering 1973-1983 research that meets the inclusion criteria described below, emphasizing findings that conflict or seem counterintuitive over findings that seem obvious and clear cut. Where findings conflict, we seek to identify methodological or contextual (subject matter, grade level, etc.) factors that may explain apparent contradictions. In this regard, the chapter builds upon reviews and methodological commentaries published by Berliner (1976,1977,1979), Borich and Fenton (1977), Brophy (1979), Brophy and Evertson (1978), Centra and Potter (1980), Cruickshank (1976), Denham and Lieberman (1980), Doyle (1977), Flanders and Simon (1969) Gage (1978,1983), Good (1979), Good, Biddle, and Brophy (1975), Heath and Neilson (1974), Kyriacou and Newson (1982), Medley (1979), Peterson and Walberg (1979), Rosenshine (1976,1979,1983), Rosenshine and Berliner (1978), and Rosenshine and Stevens (in press).

Following this introduction, the paper briefly reviews progress prior to 1970, describes *zeitgeist* trends and methodological improvements that led to the large field studies of the 1970s, details these studies and their findings, integrates these data with other data linking teacher behavior to student achievement, assesses the power and limits of the data, and discusses current trends and probable future directions.

Criteria for Inclusion

We focus on research able to be generalized to typical elementary and secondary school settings, using the following criteria.

1. Focus on normal school settings with normal populations. Exclude studies conducted in laboratories, industry, the armed forces, or special facilities for special populations.

2. Focus on the teacher as the means of instruction. Exclude studies of programmed instruction, media, text construction, and the like.

3. Focus on process-product relationships between teacher behavior and student achievement. Discuss presage and context variables that qualify or interact with process-product linkages, but exclude extended discussion of presage-process or context-process research.

4. Focus on measured achievement gain, controlled for entry level. Discuss affective or other outcomes measured in addition to achievement gain, but exclude studies that did not measure achievement gain or that failed to control or adjust for students' entering ability or achievement levels.

5. Focus on measurement of teacher behavior by trained observers, preferably using low-inference coding systems. Exclude studies restricted to teacher self-reports or global ratings by students, principals, and so on, and experiments that did not monitor implementation of treatment.

6. Focus on studies that sampled from well described, reasonably coherent populations. Exclude case studies of single classrooms and studies with little control over or description of grade level, subject matter, student populations, and so on.

7. Focus on results reported (separately) for specific teacher behaviors or clearly interpretable factor scores. Exclude data reported only in terms of typologies or unwieldy factors or clusters that combine disparate elements so as to mask specific process-outcome relationships, or data reported only in terms of general systems of teacher behavior (open vs. traditional education, mastery learning, IPI, IGE, etc.).

Overlap With Other Chapters

Some studies that meet the above criteria are treated briefly or excluded because they are covered elsewhere in the *Handbook of Research on Teaching*. To avoid unnecessary overlap with other chapters, we adopted the following criteria.

1. Focus on elementary and secondary classrooms. Exclude research in preprimary and post-secondary classrooms.

2. Focus on the teacher or class as the unit of analysis (teacher effects). Exclude studies in which the principal, school, or curriculum is the unit of analysis, or in which individual students or subgroups within classes are being compared (Aptitude-Treatment Interaction studies).

3. Focus on classroom management correlates of achievement outcomes, but minimize discussion of the details of effective classroom management (see *Handbook*, Chapter 16).

4. Focus on teacher behaviors that appear to apply to several subject matter areas. Exclude research on teacher behavior so subject-specific as to be more appropriate for Chapters 33-39 in the *Handbook of Research on Teaching*.

5. Focus on teachers working in naturalistic settings under ordinary conditions. Exclude studies of teachers trained to implement elaborately developed instructional systems (See *Handbook*, Chapter 15).

6. Focus on substantive findings. Discuss observational methods and statistical analyses to the extent necessary to clarify the data, but minimize general discussion of the relative merits of different observation approaches, raw versus standardized scores, regression versus correlation, and so on.

Although exclusive in many respects, these criteria still define a broad range of research as relevant to this chapter--most studies in which objectively measured teacher behavior was linked to adjusted achievement by elementary or secondary students. Few such studies have been done, however. Using similar but looser criteria, Rosenshine (1971) located only about 50 studies linking teacher behavior to student achievement (of these, less than 30 meet our criteria). More recently, Medley (1977,1979), using similar but more stringent criteria, excluded all but 14 studies (he only discussed correlations of .39 or higher). Thus, despite the importance of the topic, there has been remarkably litle systematic research linking teacher behavior to student achievement.

A major reason for this is cost. Classroom observation is expensive. Except for a brief period in the 1970s when the National Institute of

Education was able to fund several large field studies, investigators have not had the resources needed to do process-product studies that involve both large enough samples to allow the use of inferential statistics in analyzing the data and extensive enough observation in each classroom to allow comprehensive and reliable sampling of teacher behavior.

Historical Overview of the Field

In addition to cost, historical influences on the conceptualization and measurement of teacher effectiveness that guided research on teaching slowed development of the field. Medley (1979) has identified five successive conceptions of the effective teacher: (1) possessor of desirable personal traits, (2) user of effective methods, (3) creator of a good classroom atmosphere, (4) master of a repertoire of competencies, and (5) professional decision maker who has not only mastered needed competencies but learned when to apply them and how to orchestrate them.

Early concern with teachers' personal traits led to presage-product rather than process-product studies. Presage variables included such teacher traits as appearance, intelligence, leadership, and enthusiasm. "Product" variables were usually global ratings by supervisors or principals. This approach produced some consensus on virtues considered desirable in teachers, but no information on linkages between specific teacher behaviors and measured student achievement.

The subsequent methods focus produced experiments comparing the measured achievement of classes taught by one method with that of classes taught by another. Unfortunately, however, the majority of these studies produced inconclusive results because the differences between methods were not significant enough to produce meaningful differences in student achievement (Medley, 1979). Furthermore, the significant differences that did appear tended to contradict one another. Finally, almost all of these studies included only a few classes and inappropriately used the student rather than the class as the unit of analysis; thus effects due to methods were confounded with whatever other differences existed between the teachers (for treatments administered to intact classes, data should be aggregated and analyzed at the level of class means, and degrees of freedom should be calculated on the basis of the number of classes--not the total number of students--observed). Because of these and other difficulties, reviewers such as Morsh and Wilder (1954) and Medley and Mitzel (1963) concluded that efforts to identify effective teaching had not paid off, and that no specific teacher behavior had been linked unequivocally to student achievement.

The 1950s and 1960s brought concern about creating a good classroom climate and about the teaching competencies involved in producing student achievement. This led to an emphasis on measurement of teacher behavior through systematic observation, and to a proliferation of classroom observation systems. Some reviewers, encouraged by this progress, noted that improved process-product results could be expected if these advances in objective measurement of teacher behavior could be linked with objective measurement of student achievement. In fact, Gage (1965) and Flanders and Simon (1969) were able to report modest progress.

Other reviewers, however, were prepared to give up on this line of research, and many salient events of the 1960s and early 1970s appeared to support their point of view. One important trend was an emphasis on the curriculum over the teacher. In contrast to the research on teacher effects, studies of curriculum effects usually produced clear results indicating that students learned the content to which they were exposed (Walker & Schaffarzick, 1974). Although such curriculum-effects research is silent on the question of teacher effects, it was sometimes taken to imply that teacher effects are unimportant. Furthermore, most of the highly publicized post-Sputnik federal initiatives in education concerned curriculum reform rather than teacher training. To the extent that developers considered how (not just what) to teach, they made prescriptions based on intuition or ideology rather than objective data. They seldom felt the need to experiment with ways of teaching the content, and either trained teachers to perform according to prescribed patterns or tried to develop teacher-proof curricula that would deliver the content to the students directly rather than depend on teachers to do so.

Early school-effects research also minimized the apparent contributions of teachers. In particular, interpretations of the Coleman report (Coleman et al., 1966) and its reanalyses by Mosteller and Moynihan (1972) and by Jencks et al., 1972) seemed to indicate that teachers did not have important differential effects on student achievement. This conclusion received much more publicity than did criticisms indicating, among other things, that the study did not include systematic observation of teacher behavior and that it precluded the possibility of assessing individual teacher effects because it used the school rather than the teacher as the unit of analysis (Good et al., 1975).

Rosenshine (1970a) questioned the stability of teacher behaviors observed in process-product studies, noting that the few stability coefficients that had been reported were rather low. This called into question the meaningfulness of even low inference measures of teacher behavior (What is the value of improving measurement if the teacher behavior being measured is not stable?). Finally, Popham (1971) failed to find systematic differences in teacher behavior between trained instructors and comparison instructors who lacked special training, leading him to question whether teachers have any special expertise at all.

Yet, despite all this, significant progress occurred in the 1960s. Convinced of the validity of the process-product approach, Biddle, Gage, Medley, Soar, and others made important conceptual and methodological advances. Meanwhile, Bellack, Flanders, Hughes, Taba, and others contributed new observation systems and created interest in new process variables. By 1970, there were more than 100 classroom observation systems (Simon & Boyer, 1967, 1970). Many had been developed originally for teacher training rather than research purposes. In fact, most of the guidelines for using these systems to observe and give feedback to teachers were based on ideological commitments, and some even were contradicted by existing data (Rosenshine, 1971; Dunkin & Biddle, 1974). However, once in existence, these measurement devices and related concepts provided new tools for new process-product research.

Observation systems gradually became more sophisticated and comprehensive, especially in measuring teacher behavior related to the cognitive objectives of instruction (earlier emphasis had been mostly on affective aspects). Problems connected with reliabilities of the behaviors being measured proved solvable, at least to a degree, through increasing the amounts of observation time allocated per classroom and instituting better controls over the contexts within which observations were scheduled. Studies using the class as the unit of analysis began to show significant, and sometimes stable, teacher effects and process-product linkages.

Rosenshine (1971) reported that data from different investigators using different methods indicated that certain teacher behaviors were consistently correlated with student achievement gain. These correlations were not always significant, and typically were only marginal to moderate in strength even when they did reach significance. Nevertheless, the consistency in findings for certain variables was encouraging. Strong criticism of students was correlated negatively with achievement gain (mere negation of incorrect responses was unrelated or correlated positively). Positive correlates included warmth, businesslike orientation, enthusiasm, organization, variety in materials and academic activities, and high frequencies of clarity, structuring comments, probing questions asked as follow up to initial questions, and focus on academic activities. No significant correlations were found for nonverbal expression of approval, use of student ideas, or amount of teacher talk. Mixed results were reported for verbal praise, level of difficulty of instruction or of teacher questions, and amount of student talk. Rosenshine suggested that the latter variables might show inverted-U curvilinear relationships to student learning or might interact with students' individual differences.

Rosenshine's review helped pull together and define the field, and it drew attention to some important methodological and interpretive issues. Besides noting that teacher variables might have non-linear relationships to student achievement or might interact with students' individual differences, Rosenshine stressed the need to consider context or sequence factors that might affect the meanings of teacher behavior. He noted, for example, that frequency counts of teacher approval or criticism are not very useful without information about the contexts within which these teacher evaluations were delivered. Similarly, the usefulness of high- versus low-level teacher questions might be expected to vary with subject matter and grade level, so that box scores summarizing results across all studies might yield puzzling contradictions, but analyses of findings within comparable contexts might yield regularities. Finally, Rosenshine noted that qualitative distinctions in coding related but different teacher behaviors (mere feedback vs. praise or blame, brief vs. extended use of student ideas) produced more coherent results than coding with less finely differentiated categories.

Besides documenting progress, the Rosenshine (1971) review illustrated the interpretive dilemmas involved in trying to integrate and explain process-product findings. Sometimes investigators use different terminology but measure similar teacher behaviors and produce comparable findings, and sometimes they use similar terminology but measure quite different teacher behaviors and produce findings that are unrelated. If data are reported only for combination scores composed of disparate elements, it is impossible to determine whether a correlation involving the combination score holds for any particular element individually. In fact, as Rosenshine (1971) noted, different items grouped in combination scores for theoretical reasons may have contrasting patterns of correlation with achievement.

Even where clear data link reasonably specific teacher behaviors to student achievement, the causal linkages underlying the correlation remain unknown pending follow up experimentation. For example, what is one to make of the negative relationship between frequency of severe criticism and student achievement gain? Strong teacher criticism of students rarely occurs (the correlations obtained for this variable represent the difference between teachers who seldom criticize and those who rarely or never criticize). It seems likely, then, that the correlation is not so much due to a direct negative effect of teacher criticism on student learning as to a tendency for teacher criticism to be associated with other teacher characteristics that affect student learning more directly. Perhaps criticism is more frequent among poor classroom managers who are often frustrated by student disruptions, for example, or among poor instructors who are often frustrated by student failure.

Researchers have attempted to solve these interpretive dilemmas with varying success. Logical clustering, factor analysis, and related methods are often used for reducing the data, but these procedures will mask rather than illuminate process-product relationships if the resulting scores combine teacher behaviors that should be kept separate. We believe that analyses of process-product data should focus on identifying and coming to understand the reasons for reliable relationships. Data reduction techniques can help accomplish this when the measures being combined are aspects of the same basic teacher behavior, but otherwise, correlational patterns should be examined separately for each measure.

Coming to understand process-product data requires attention not only to correlation coefficients, but also to the means and patterns of variation in the teacher behaviors involved (as in the above example involving teacher criticism) and to context factors (grade level, subject matter, etc.) that may qualify generalization of findings. Most reviewers have tried to deal with these complexities by identifying variables studied similarly in different studies and describing general trends in the findings, perhaps adding qualifications based on context variables as well. Dunkin and Biddle (1974) formalized this approach by constructing boxes that concisely summarized the existing research on various teacher behaviors. More recently, this general approach has been formalized still further in meta-analysis procedures developed by Glass and Smith (1978).

We have taken a different approach in this chapter. Rather than organize according to teacher behavior variables and compute box scores or meta-analyses that would largely repeat ground covered earlier by Dunkin and Biddle, Medley, Rosenshine, and others, we have decided to organize the review around what appear to be the major programmatic studies in the field, and use their common findings to induce and integrate generalities. In contrast to the box score and meta-analysis approaches, this approach focuses on the studies that seem most likely to produce valid and generalizable findings, and takes into consideration grade level, subject matter, type of teacher and classroom, amount and type of measurement of teacher behavior, and other factors unique to specific studies that may be useful in interpreting their findings. It involves more judgment and less mathematical precision than the other approaches, but we believe that it is better suited to the task of coming to understand the reasons for observed process-product relationships (and especially for resolving apparent discrepancies and explaining real discrepancies in the findings).

Progress in the 1970s

Several events occurring in the early 1970s helped to consolidate the progress of the 1960s and prepare the way for subsequent developments. One was the publication of a chapter by Rosenshine and Furst (1973) in the *Second Handbook of Research on Teaching* on the use of direct observation to study teaching. These authors noted that consistent findings had begun to accumulate and discussed the relative merits and potential research uses of the classroom observation instruments that had accumulated and been catalogued in *Mirrors for Behavior* (Simon & Boyer, 1967, 1970). They also called for programmatic work on the "descriptive-correlational-experimental loop," in which classroom observation would lead to the development of instruments to measure (describe) teaching in a quantitative manner. Next, correlational studies would be conducted to relate the descriptive variables to achievement, and, finally, experimental studies would be conducted to test promising correlational relationships for causal effects.

Rosenshine and Furst also made methodological suggestions that foreshadowed later developments: (1) attend to the cognitive (rather than affective) aspects of teaching, because these are the ones most likely to determine learning; (2) insure that tests reflect the content taught; (3) use more complex and varied coding systems; (4) attend to sequences of events; (5) tailor the observation system to the subject matter and context; (6) sample behavior that is representative of the teachers' typical patterns; and (7) develop a rich bank of process-process and process-product data in each study to facilitate interpretation of the findings.

In 1974, Dunkin and Biddle published *The Study of Teaching*, which reviewed and critiqued all extant research that included low inference measurement of teacher behavior. This book helped define the field of research on teaching and differentiate it from other forms of educational research. Following Mitzel (1960), Dunkin and Biddle organized the research into a model, featuring presage, process, product, and context variables, and constructed boxes summarizing what was known about the frequencies of various teacher behaviors and about their relationships to context, presage, product, and other process variables. They complained of the widespread tendency to make educational prescriptions based on untested theoretical commitments rather than convincing empirical data, stating that before attempting to implement a research finding in the schools, one would want to know:

> that the concepts used in the finding are meaningful, and that they had been measured with instruments that were valid and reliable; that the studies reporting the finding had used valid, uncontaminated designs; that the effect claimed was strong, that it was independent of other effects, and that the independent variable claimed for it was truly independent; that the effect applied over a wide range of teaching contexts, or if not, to what range it was limited; and finally that we understood why the effect took place. (p. 358)

At the time, most progress had taken place with regard to the first two of these concerns. This is still true, although progress in the latter three areas has also occurred in recent years, and we intend to give particular emphasis to these concerns here (especially the last two; in regard to the third, we are not so much concerned about the strength or independence of process-product relationships as we are about describing and explaining them--whether they are weak or strong, linear or nonlinear, independent or nested within larger patterns).

Dunkin and Biddle emphasized the need to attend to context variables--both to include them in the design or a least control them in selecting the teacher sample and the activities to be observed and to suggest limits on the generalization of results. They also chided researchers for fundamental yet common mistakes (failure to sample adequately, inappropriate use of inferential statistics, failure to report basic descriptive data) and called for more comprehensive investigations designed to develop theory and explain findings rather than merely to garner support for some pet idea.

Another major factor influencing progress in the 1970s was the involvement of federal agencies, particularly the Office of Education (OE) and the National Institute of Education (NIE). In particular, the OE's funding of evaluation studies of Project Follow Through and the NIE's funding of several large-scale field studies and (later) experiments allowed investigators to conduct process-product research on a scale never approached previously. Furthermore, the NIE convened a national conference on studies in teaching in 1974, bringing together leaders in the field to assess progress, identify needed methodological improvements, and suggest research priorities. Later, the NIE followed up by establishing the Invisible College for Research on Teaching, an informal organization of classroom researchers who gather prior to the annual American Educational Research Association meetings to share state of the art information. Both the agenda setting at the 1974 conference and the subsequent Invisible College activities helped pull together and unify process-product research specifically and research on teaching generally as viable fields of scientific inquiry. More recently, the NIE sponsored a conference to review research on teaching and summarize its implication for practitioners. The papers were later published in the March 1983 issue of the *Elementary School Journal*.

The report of Panel 2 of the 1974 conference (National Institute of Education, 1974) produced a list of key methodological considerations for process-product researchers, identifying the following as desirable: programmatic, cumulative research designs; letting the goals of the project, and not habit or convenience, determine what and how to measure; multiple measurement of a variety of outcomes (product variables); considering non-linear process-product relationships; considering complex interactions among variables (suppressor effects, moderator effects, etc.); eliminating or controlling entry level differences in student ability or achievement; including both high and low inference measures of a variety of process behaviors; selecting samples of teachers and classrooms to insure comparability and representativeness; collecting enough data in each classroom to insure reliability and validity (or, alternatively, controlling classroom events by standardizing lessons and materials); controlling for Hawthorne effects and monitoring implementation in experimental studies; insuring adequate variance and stability in relevant teacher behaviors in naturalistic studies; taking into account patterns of initiation and sequence in teacher-student interaction; and devising scoring systems that allow for more direct comparison of teachers or students than mere frequency counts provide (for example, teachers can be compared more validly using the percentages of their students' correct answers that are praised than using the rates of such praise, because percentage scores take into account differences in frequency of correct student answers).

Major Programs of Process-Product Research

No study has yet been done that includes all of these desirable characteristics, but the process-product research of the 1970s came much closer to approaching these ideals than earlier research had done and, correspondingly, yielded more satisfactory results. We now turn to these findings, starting with the work of research teams who studied process-product questions programmatically in series of related studies.

Canterbury Studies

A series of studies done at the University of Canterbury in New Zealand began with a correlational study by Wright and Nuthall (1970), in which teachers taught science lessons to groups of 20 randomly selected third graders. There were no significant correlations (with achievement adjusted for IQ and general science knowledge) for total teacher or pupil talk, total teacher structuring comments, percentage of structuring that occurred immediately following questions, or starting lessons with reviews of the previous lesson; positive relationships for percentage of structuring that occurred at the ends of episodes initiated by questions, percentage of closed (rather than open) questions, praising or thanking students for their responses, asking single questions rather than two or more questions in series, and concluding lessons with reviews; and a negative relationship for student failure to respond to questions.

Redirection of the same question to another pupil following the response of the first pupil correlated positively with achievement, but there were no significant relationships with elaborating or trying to elicit improvement on the original response. These measures were not coded separately for whether or not the original question was answered correctly, however, so their meanings are not clear.

Follow up studies by Hughes (1973) involved experimental manipulation of pupil participation and teacher reactions to pupils' responses during lessons taught to seventh graders about animals. The first study involved three pupil participation treatments: random response (questions addressed to students at random), systematic response (questions addressed according to pupils' seating positions), and self-selected response (questions directed only to volunteers). The results showed no differences between treatment groups and no relationship between student rate of response (whether voluntary or involuntary) and adjusted achievement.

A second study involved a more extreme manipulation, in which a randomly selected half of the students in each class were asked all of the questions, while the other half were given no chances to repond at all. Once again, however, overt participation was unrelated to achievement.

A third study dealt with teacher reactions to student response. Pupils in the "reacting" group were given frequent praise for correct answers and support, along with occasional urging or mild reproach when they failed to respond correctly. Pupils in the "no reacting" group generally received little more than a statement of the correct answer. The reacting group outgained the no reacting group, both on items related to questions asked during the lesson and on other items. Taken together, Hughes's data suggest that, by seventh grade, pupils can learn effectively without overt participation in lessons, but that their learning can be affected by teachers' reactions to the responses of the students who do participate. These teacher reaction effects appear to have been motivational (mediated by the enthusiasm and teacher demands communicated in the reacting group treatment) rather than instructional (the reacting treatment did not involve greater opportunity to participate or get information).

Nuthall and Church (1973) describe other work done at Canterbury. In one study, teachers were asked to concentrate either on teaching conceptual knowledge or on maximizing achievement test scores. The teachers intending to teach conceptual knowledge used more open-ended questions and included more logical connectives, but did less lecturing. However, these differences were unrelated to pupil test scores, either for factual knowledge or for higher level conceptual knowledge.

Another study (about teaching science concepts to 10-year-olds) involved manipulating both content coverage (how much content was introduced, to what degree of redundancy, and with how much time spent teaching it) and teacher behavior (questioning vs. lecturing). Content coverage was much more closely related to achievement. With coverage held constant, there was no difference in effects on achievement between the questioning method and the lecture method. Within the questioning method, however, contrary to Hughes's findings for seventh graders, Nuthall and Church found that students who were called on to respond learned more than those who were not.

Taken together, the Canterbury studies suggest that (1) content coverage determines achievement more directly than the particular teacher behaviors used to teach the content; (2) younger students need to participate overtly in recitations and discussions, but older ones may not require such active participation; (3) questions should be asked one at a time, be clear, and be appropriate in level of difficulty so that students can understand them (most such questions will be lower order); (4) teacher reactions to student response that communicate enthusiasm for the content and support (or if necessary, occasional teacher demands) on the students are more motivating than matter-of-fact reactions; and (5) teacher structuring of the content, particularly in the form of reviews summarizing lesson segments, is helpful.

Flanders

Perhaps the most useful programmatic process-product research conducted prior to the 1970s was the work of Ned Flanders and his associates (Flanders, 1970), using the Flanders Interaction Analysis Categories (FIAC). Flanders believed that there was too much teacher talk and not enough student talk in most classrooms and that teachers should be more indirect--should do more questioning and less lecturing and, in particular, should more often accept, praise, and make instructional use of the ideas and feelings expressed by their students. Flanders was interested primarily in the effects of teacher indirectness on student attitudes (liking for the teacher and the class), but he also included measures of adjusted student achievement in five studies conducted between 1959 and 1967.

The basic procedures were as follows: first, pupil attitude inventories were administered, and classes located at the extremes of the distribution of pupil attitudes were selected for further study (sometimes other classes were also included). Then, entering achievement level was assessed, and the classes observed with FIAC. The teachers worked in their regular classrooms with their regular students during these observations, but were observed teaching specially prepared experimental teaching units (similar to regular units but on different topics). This minimized the degree to which mastery of the content taught would be affected by previous school learning. Coders would observe classroom interaction for three seconds, then code the interaction into one of the 10 FIAC categories (shown in Table 1), then observe for another three seconds. The raw data were summed to produce frequency scores, which in turn were added to produce combination scores or divided to produce ratio scores (see Table 2). Flanders was most interested in the ratio of indirect to direct teaching. In his earlier work, he classified lecturing,

Table 1

Representative Data for Various Types of Junior High School Classrooms Described in Terms of the Flanders' Interaction Analysis Categories (FIAC), Expressed as Percentages of Total Interactions Observed

Types of Teacher Behavior		Types of Classrooms				Total
		Math Indirect	Math Direct	Social Studies Indirect	Social Studies Direct	
Indirect	1. Accepts feeling	.23	.11	.11	.03	.12
	2. Praises, encourages	1.69	1.06	1.25	1.14	1.28
	3. Uses pupil ideas	8.11	2.63	8.28	3.03	5.51
	Indirect subtotal	10.03	3.80	9.64	4.20	6.91
	4. Asks questions	12.52	9.53	10.75	10.80	10.90
	5. Lectures	46.72	40.83	37.45	25.67	37.67
Direct	6. Gives directions	3.38	8.64	4.29	9.86	6.54
	7. Criticizes, justifies authority	.94	4.66	1.69	5.32	3.15
	Direct subtotal	4.32	13.30	5.98	15.18	9.69
	8. Pupil talk, response	10.73	13.02	17.54	21.49	15.70
	9. Pupil talk, initiate	6.12	6.74	9.48	8.70	7.76
	10. Silence, confusion	9.56	12.79	9.16	13.94	11.36
No. of classrooms		7	9	7	8	31
No. of interactions observed		26,083	32,726	28,194	23,641	110,644

Source: Flanders, *Teacher Influence, Pupil Attitudes, and Achievement* (Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1965), pp. 75-76.

Table 2

Correlations Between Flanders' Teacher Behavior Variables and Student Adjusted Achievement and Attitudes in Five Studies

Variable	Computation Rule	Correlations with Adjusted Achievement — Study/Grade Level 2nd	4th	6th	7th	8th	Correlations with Class Attitude — Study/Grade Level 2nd	4th	6th	7th	8th
1. Indirectness Proportion (i/i+d)	Sum of Accepts Feeling (1) + Praise (2) + Uses Pupil Ideas (3) codes divided by Sum of Accepts Feeling (1) + Praise (2) + Uses Pupil Ideas (3) + Gives Directions (6) + Criticizes or Justifies Authority (7) codes.	-.07	.31	.22	.48*	.43*	.13	.64*	.49*	.34	.58*
2. Sustained Acceptance Sum	Sum of Uses Pupil Ideas (3) codes which were followed by another Uses Pupil Ideas (3) code.	-.45	.19	.30	.40	.19	.13	.52*	.40*	.33	.31
3. Indirectness Sum	Sum of Accepts Feeling (1) + Praise (2) + Uses Pupil Ideas (3) + Asks Questions (4) codes.	.05	-.08	.26	.25	.45*	.45*	.34	.40*	.16	.51*
4. Questions Sum	Sum of Asks Questions (4) codes	.07	-.19	.11	-.06	.44*	.49*	-.06	.27	.00	.47*
5. Teacher Talk Sum	Sum of Codes in Categories 1-7.	.30	.08	.11	.02	.45*	.38	.10	.24	.15	.61*
6. Restrictiveness Sum	Sum of Gives Directions (6) + Criticizes or Justifies Authority (7) codes.	-.10	-.24	-.04	-.61*	-.34	-.09	-.17	-.37	-.43	-.66*
7. Restrictive Feedback Sum	Sum of Pupil Response (8) + Pupil Initiation (9) codes which were followed by Gives Directions (6) or Criticizes or Justifies Authority (7) codes.	.18	-.34	-.32*	-.50*	-.43*	.02	-.32	-.29	-.47*	-.62*
8. Negative Authority Sum	Sum of (6) codes followed by (7) codes + Sum of (7) codes followed by (6) codes.	.05	-.23	-.15	-.62*	-.25	-.22	-.22	-.32*	-.43	-.59*
9. Praise Sum	Sum of Praise (2) codes	.25	-.13	.36*	-.23	.30	.08	.40	.35*	-.34	.38
10. Flexibility	The i/d ratio is computed separately for each classroom observation (Sum of 1 + 2 + 3 divided by sum of 1 + 2 + 3 + 6 + 7). Then, the lowest of these ratios is subtracted from the highest to obtain the range.	-.07	.46*	.19	.37	.43*	.12	.08	.41*	.13	.43*
	Number of classes	15	16	30	15	16	15	16	30	15	16

*$p < .05$

(Constructed from data given on pp. 394-305 of Ned A. Flanders, Analyzing Teacher Behavior, Reading, Mass., Addison-Wesley, 1970).

giving directions, criticizing, and justifying authority as direct influence techniques, and asking questions, accepting and clarifying ideas or feelings, and praising or encouraging as indirect techniques. Later he eliminated lecturing and questioning from his scoring of direct and indirect teaching.

In *Analyzing Teacher Behavior*, Flanders (1970) reviewed his own work and that of others who had used FIAC to link teacher-student interaction to student attitudes or achievement. Representative data from five of his own studies are shown in Table 2. Several facts about these data are noteworthy. First, they do not support the notion that teachers talk too much. In all five studies, teacher talk correlated *positively* with both achievement and attitude. Thus, although about two-thirds of the talk in classrooms is teacher talk, there is no reason to believe that such talk is inappropriate or that it indicates that teachers are oppressive, unduly dominant, and the like.

Second, the data generally support Flanders' hypotheses (more for attitude than for achievement), although the second grade data are systematically less supportive than the data from the other four studies. Correlations with indirectness, praise, and acceptance of student ideas tend to be positive, and correlations with restrictiveness and negative authority tend to be negative.

Third, the negative correlations for restrictiveness and criticism tend to be stronger and more consistent than the positive correlations for praise and acceptance of student ideas (especially in the data for student achievement). Furthermore, although praise and sustained acceptance are lumped together in computing indirectness scores, these teacher behaviors often correlate in opposite directions with student achievement.

Finally, the flexibility score generally correlates positively with student attitude and achievement, indicating the need to tailor techniques to the situation rather than trying to maximize indirectness at all times. Following Soar (1968), Flanders (1970) noted that teacher behavior variables may have "inverted U" curvilinear relationships or other nonlinear relationships with student achievement, so what is optimal teacher behavior may vary with the situation. He suggested that lower levels of indirectness might be appropriate for factual or skill learning tasks and higher levels for tasks involving abstract reasoning or creativity. We agree with these observations and believe that they help explain the discrepant second grade data. Because most school activities in the primary grades involve low level factual and skill learning, there is less reason to expect indirectness variables to relate to achievement in these grades in the same ways they do at higher grades.

In summary, except for the second grade data, the data shown in Table 2 suggest positive relationships between indirect teaching and achievement (although we have direct data only for sustained acceptance and praise; separate correlations are not given for accepting students' feelings, using student ideas, giving directions, or criticizing or justifying authority). Should one conclude, then, that students beyond the primary grades will achieve more if their teachers become more indirect? We think not, for several reasons.

The first, of course, is that the data are correlational. We could just as well conclude that student achievement causes teacher indirectness or that both variables covary with some more fundamental but unmeasured third factor. Furthermore, several experimental studies comparing indirect to direct teaching failed to produce significant group differences in achievement (Rosenshine, 1970b). Thus, even when correlated with achievement, teacher indirectness variables do not necessarily cause it.

Second, as noted by Flanders (1970) himself and elaborated by Barr and Dreeben (1978), the teacher behaviors included in indirectness ratios only apply during recitations and other activities in which the teacher is instructing the whole class or a significant subgroup, and furthermore apply to only a small proportion of the interaction that occurs in these settings. The data in Table 1, from mathematics and social studies classes, are typical. Note that only about 7% of the codes are classified as indirect and only about 10% as direct. Compare this with about 11% for teacher questions, 38% for lecturing, and 23% for pupil talk. Teacher indirectness behaviors subsume only a minority of classroom events and have nothing directly to do with the quantity or quality of instruction in subject matter content. Furthermore, teachers that use an indirect style provide only 5-6% more indirect teaching than do direct-style teachers, but yet provide about 9% more lecturing. It is possible that this, rather than indirectness, explains the differences in achievement (Flanders did not provide correlations specific to teacher lecturing; the teacher talk variable includes all seven of the teacher categories).

Third, note that indirectness behaviors occur in public settings in which the teacher is presenting information, conducting a recitation or drill, or leading a discussion. It may be that teachers using an indirect approach elicit more achievement not so much because they are more likely to use indirect methods during group instruction, but because they do more group instruction in the first place (group instruction maximizes opportunities to accept students' feelings, praise, or use their ideas, and minimizes the need to give directions or criticize). Indirect teachers may actively instruct their students more often than teachers using a direct style.

A related point is that the FIAC system requires that every three-second observation be coded, so that procedural and conduct interactions get mixed in with academic interactions instead of being coded separately or ignored. As a result, several FIAC categories, especially six and seven, include significant proportions of codes based on nonacademic interaction (many of teachers' directions are procedural, and most of their criticism is for misconduct rather than incorrect answers). Teachers who frequently give procedural directions or behavioral criticism usually do so because their students are often confused, off task, or disruptive. Thus, the FIAC system has a built in tendency to classify as direct those teachers who students spend less classroom time engaged in academic tasks.

Finally, the FIAC system did not distinguish between simple affirmative feedback and praise nor between simple negation and criticism. Consequently, to the extent that statements coded as praise or criticism did refer to academic responses, the majority merely affirmed or negated the correctness of the student's statement. Also, the measures used were simply the summed frequencies of the categories praise and criticism (rather than the percentages of correct answers praised and wrong answers criticized)--measures that depended in large part on how frequently the students in a class gave correct answers. In turn, this depended on pupil ability and comprehension of the material as well as on the teachers' skill in presenting the material and posing clear and appropriate questions. Thus teachers' content presentation and questioning skills may have affected their indirectness scores.

These methodological and interpretive comments are included here not so much to criticize Flanders' work (he advanced the field and was ahead of his time in many ways) as to clarify its interpretation and its relationships to subsequent work by others. At first, Flanders' data seem to contradict some of the most common findings (reviewed below) of the 1970s. However, Flanders' data are seen to be compatible with these later findings when it is recognized that teacher lecturing is not included in those measures of direct teaching that correlate negatively with achievement; relationships are curvilinear, revealing a lower optimum amount of indirectness in basic skills lessons; levels of student ability and motivation will affect the indirectness scores attributed to teachers, and teachers who spend more time actively instructing their students and less time dealing with procedural or student conduct concerns are likely to get higher indirectness scores.

Soar and Soar

As noted above, the theorizing of Robert Soar (1968) concerning inverted-U curvilinear process-outcome relationships is useful in interpreting the Flanders (1970) data. Soar also conducted five process-outcome studies in the 1960s and 1970s, several in collaboration with Ruth Soar. These studies typically involved multiple measurement of student entry characteristics in the fall, of classroom processes in the middle of the school year (typically based on four to eight half-hour visits per class), and of student outcomes in the spring. The sample descriptions and references for these five studies are: (1) 55 urban classrooms, grades 3-6, all white and predominantly middle and upper socio-economic status (SES) (Soar, 1966); (2) 20 first-grade classrooms in Project Follow Through, mixed racially but with predominantly low SES pupils (Soar & Soar, 1972); (3) 59 fifth-grade classrooms, mixed racially but with predominantly low SES pupils (Soar & Soar, 1973,1978); (4) 22 urban, first-grade classrooms, mixed racially and heterogeneous in SES (Soar & Soar, 1973, 1978); (5) 289 Follow through and comparison classrooms in the primary grades, predominantly low in SES (Soar, 1973).

Two observation systems were used in the first study, one an elaboration of FIAC and one concerned with nonverbal behavior and expression of affect.

The other studies used four systems, two coded on-the-spot and two coded later from audiotapes. The first looked at classroom management, pupil response to it, and the teacher's and pupils' expression of affect. The second categorized the teacher's development of subject matter, using concepts from Dewey's experimentalism. The third characterized the cognitive level of discourse, using Bloom's taxonomy of cognitive objectives. Finally, the fourth system was the elaboration of FIAC.

Although combinations of factor analysis and rational cluster analysis were used to reduce the process data, the resultant factors usually possessed conceptual clarity and face validity as measures of specific teacher behavior. Factor scores were then entered into analyses designed to reveal both linear and nonlinear relationships with achievement, which was adjusted not only for entry level but frequently for personal characteristics such as dependency, anxiety, or cognitive style as well. The Soars (Soar, 1977; Soar & Soar, 1979) have integrated findings from the first four of the studies listed above, using some key conceptual distinctions.

Conceptual distinctions. The first distinction is between *emotional climate* factors (positive or negative affect exhibited by teachers and students) and teacher *management* (or control) factors. These factors are independent: Highly controlling teachers are not necessarily rejecting or otherwise negative, and teachers who exert minimal control over pupil behavior are not necessarily student oriented or otherwise positive in their affect.

Within the sphere of emotional climate, the teacher's affect must be distinguished from the pupils' affect. Positive affect in the teacher does not necessarily imply positive affect in the students, or vice versa. Within the teacher management sphere, it is important to distinguish between *control* *of pupil behavior* (physical movement, opportunity to socialize), *control of learning tasks* (what learning tasks are selected and how are they carried out), and *control of thinking processes* (degree to which pupils are allowed or encouraged to confront the subject matter at a variety of cognitive levels or to pursue divergent ideas). Here too, there are no necessary relationships. A teacher who highly controls physical movement and nonacademic behavior might or might not allow considerable pupil choice of learning activities or opportunity to engage in a variety of thinking processes.

Finally, the Soars also note that teacher control can be exercised either by establishing rules and routines ("established structure"), or by issuing directives, asking questions, or otherwise structuring pupil response through immediate face-to-face interaction ("current interaction"). Once again, these elements are independent: Teachers who control through established structure may or may not highly control their daily interactions with the students.

Emotional climate. The Soars draw several conclusions that not only make good sense and fit the data from their own four studies, but also fit data from other investigators. First, there is a disordinal relationship between emotional climate and achievement gain. Negative emotional climate indicators (teacher criticism, teacher or pupil negative affect, pupil resistance) usually show significant negative correlations with achievement, but positive emotional climate indicators (teacher praise, positive teacher or pupil affect) usually do not show significant positive correlations. Most relationships are insignificant, and some are negative (especially in Soar's first study, where the students were from predominantly high SES backgrounds). Thus these data do not support the notion that efficient learning requires a warm emotional climate. It is true that negative climates appear dysfunctional, but neutral climates are at least as supportive of achievement as more clearly warm climates.

Teacher management. Measures of teacher control typically relate either positively or curvilinearly to achievement. Indicators of teacher control over student behavior (physical movement, socializing) show positive relationships. Students learn more in classrooms where teachers establish structures that limit pupil freedom of choice, physical movement, and disruption, and where there is relatively more teacher talk and teacher control of pupils' task behavior.

Indicators of high teacher control of learning tasks also correlate positively with achievement. This was seen regularly for measures of teacher-focused academic instruction (whole class or small group). In addition, the fifth-grade study showed positive correlations for indicators of good management of independent seatwork time (pupils were usually engaged in their work, and alternative activities were available when they finished).

This general pattern of positive linear relationships was qualified by several curvilinear relationships, however. Inverted-U relationships were seen in one study for recitation activity and in another for drill and for teacher directed (vs. pupil selected) activity. Thus, within the range of teacher control of learning tasks observed, the teachers who exerted greater control generally elicited higher achievement, but the relationship was ultimately curvilinear. Beyond an optimal level, additional teacher direction, drill, or recitation became dysfunctional (not because the extra instruction undermined existing learning, but because it was unnecessary and used up time that could have been spent moving on to new objectives).

The results for indicators of teacher control over pupil thinking varied with SES and grade level. In the study involving high SES students in grades 3-6, achievement related positively to high cognitive-level activities, and either positively or curvilinearly to indirect instruction. Codes for high cognitive level and indirectness are associated with discussion (rather than recitation or drill) activities. In contrast, achievement in the first-grade and low SES fifth-grade classes was associated with recitation or drill, with activities characterized by giving and receiving information, and by narrow rather than broad teacher questions. Taken together, the data suggest that ". . . greater amounts of high cognitive-level interaction are dysfunctional for young pupils, especially those of lower ability, but may become functional for older elementary pupils, especially those of higher ability" (Soar & Soar, 1979, p. 114).

There were also indications that the optimal level of teacher control (vs. student freedom) varied with learning objectives. Within any particular study, gains on lower level objectives were associated primarily with recitation, drill, and other low cognitive-level, high teacher-focus activities, and gains on tests of higher level skills were associated more with discussion and other activities offering more pupil freedom. Thus,

> some degree of pupil freedom, within a context of teacher involvement that maintains focus, was related to gain . . . for lower grade pupils, greater amounts of high cognitive-level interaction are not functional . . . the amount of pupil freedom that is most functional for both learning tasks and thinking depends on the complexity of the learning task--for more complex tasks, a somewhat greater degree of freedom is functional, but even then it may be too great. (Soar & Soar, 1979, pp. 117-118)

Finally, these studies indicate that student SES interacts with the findings for emotional climate and teacher control. Positive affect appears to be more functional and negative affect more dysfunctional for low SES pupils than for high SES pupils. Also, a greater degree of teacher control and structuring appears to be functional for low SES pupils than for high SES pupils. The work of Brophy and Evertson and of Good and Grouws (to be described) support similar conclusions.

The fifth study listed above (Soar, 1973), dealing with 289 Follow Through and comparison classrooms, was not included in the syntheses by Soar (1977) and by Soar and Soar (1979), but yielded generally compatible findings. That is, in these primary grade classrooms with low SES students, achievement gain was associated with teacher-structured time spent in reading and other academic activities involving drill or convergent questions. These findings are also compatible with the results of Stallings' research on Follow Through classrooms (described next).

Stallings

Research by Jane Stallings and her colleagues has included evaluation of Project Follow Through, correlational work at the third grade level, and correlational and experimental work in secondary reading instruction.

Follow Through Evaluation Study. This study (Stallings, 1975; Stallings & Kaskowitz, 1974) involved 108 first-grade and 58 third-grade classes taught by experienced teachers who were implementing one of seven Follow Through models. Each class was observed for three consecutive days, focusing on the teacher for two days and on selected students for one day. Data collection focused on events important to the program sponsors, and included details about the physical environment, data on the time spent in various activities, and frequency counts of adult-child interaction. Program models ranged from heavy emphasis on structured teaching of basic skills to open classroom approaches stressing affective objectives and self-directed learning.

The two programs with the clearest academic focus produced the strongest gains in reading and math, although the students were below average in attendance (considered a measure of student attitude toward school) and in scores on the Raven's Coloured Progressive Matrices (a test of perceptual problem solving ability administered only at the third grade level). This was one of several indications from 1970s work that the factors that maximize gain on standardized achievement tests are not necessarily the same factors that maximize progress toward other outcomes.

Implementation data indicated that most teachers followed the guidelines of their program sponsors. Consequently, as a sample, those classes contained much more variation in types of activity than would be observed in more traditional classes, as well as unusual combinations of program elements. For example, the Kansas program for the first-grade level (Ramp & Rhine, 1981) called for (1) frequent small group instruction in basic skills by a teacher, an aide, and two parent volunteers; (2) use of programmed individualized learning materials at other times; and (3) praise and tokens (backed by reinforcement menus) for good behavior and academic progress. This was the only program to use token reinforcement, and its combination of high rates of small-group instruction with high rates of individualized independent learning is unusual.

In many respects, then, the program rather than the class is the real unit for interpreting the Follow Through findings. Still, the data suggest the same general conclusions as other studies of primary grade instruction for low SES students, and in most respects, the Follow Through data are typical of data from large field studies that employ multiple measures of teacher behavior. There are a great many findings, involving more variables than classes. For example, for the 108 first-grade classes, 108 of 340 correlations were significant at the .05 level for mathematics, and 118 of 340 were significant for total reading. This clearly suggests significant process-product relationships, but the probability coefficients cannot be taken literally because the 340 process variables are neither conceptually nor statistically independent. Thus the .05 level of statistical significance is used merely as an informal guideline for interpreting the data.

The clearest and most widespread pattern involved positive correlations with achievement for process variables related to student opportunity to learn academic content (time spent in academic activities, frequencies of small or large group lessons in basic skills, and frequencies of supervised seatwork activities), and negative correlations for time spent in nonacademic activities (story, music, dance, arts and crafts) or in teacher-student interaction patterns that were not stressed in the two academic programs (particularly, open or informal patterns in which teachers mostly worked with one or two individuals rather than teaching formal lessons to groups). Almost anything connected with the classical recitation pattern of teacher questioning (particularly direct, factual questions rather than more open questions) followed by student response followed by teacher feedback correlated positively with achievement. Instruction in small groups (up to eight students) correlated positively in first grade, and instruction in large groups (nine or more students) in third grade.

In general, the major finding was that students who spent most of their time being instructed by their teachers or working independently under teacher supervision made greater gains than students who spent a lot of time in nonacademic activities or who were expected to learn largely on their own. Furthermore, although the sample was composed mostly of low SES (and thus relatively low ability) students, these main effects were elaborated by interactions with student ability: Frequent instruction by the teacher was especially important for the lowest ability students.

Compared to the findings for opportunity to learn/active instruction by the teacher, the findings for praise, criticism, and reinforcement were weaker and more mixed. Token reinforcement correlated positively with achievement in first grade, where it was used in the Kansas program, but by third grade it had been phased out. Praise for correct responses or good academic work also tended to correlate positively, but more notably in first grade than in third, for math than for reading, and for low ability students than for high ability students. Other forms of praise had mixed and mostly nonsignificant relationships. Neutral corrective feedback (involving neither praise nor criticism) usually correlated positively. Surprisingly, measures of negative corrective feedback (academic criticism) tended to correlate positively with learning gain when they did reach statistical significance (usually they didn't).

Taken together, these data on academic feedback suggest several general conclusions. (1) When teacher feedback measures are expressed as raw frequencies (i.e., number of academic praise statements observed) rather than being adjusted for frequencies and types of student academic responses (i.e., proportion of correct answers observed that were praised by the teacher), their interpretation is ambiguous. All types of academic feedback occur more often during activities in which academic responses are elicited more often in the first place (i.e., drill or recitation lessons). Therefore, a positive correlation for frequency of academic praise may occur because of a linkage between achievement and the frequency of active instruction by the teacher and not because of a more specific linkage between student achievement and teachers' tendencies to praise good academic responses when they are elicited. (2) Partly as a result, frequency measures of types of academic feedback show weaker relationships to achievement than measures of time spent in academic activities. (3) Academic praise and especially academic criticism are infrequent, and their base rates must be taken into account in interpreting their correlations with achievement. (4) Occasional praise (of perhaps 5-10% of good academic responses) tends to show weak but positive correlations with achievement, at least for younger and lower ability students. (5) Criticism for poor academic responses sometimes also shows weak positive correlations, at least by third grade, but such criticism is rare, and the operative difference is between never criticizing and criticizing only rarely. Most such criticism is for repeated inattentiveness or carelessness and thus represents an appropriate academic demand rather than an inappropriate hypercritical stance on the part of the teachers who employ it (in response to only about one percent of students' failures to respond correctly, about 0.05% of students' total academic responses). (6) These conclusions apply to academic criticism, not criticism for misconduct. The latter almost invariably correlates negatively with achievement and indicates classroom organization and management difficulties.

California ECE Study. Stallings, Cory, Fairweather, & Needels (1977) evaluated reading instruction in the California Early Childhood Education (ECE) program, which was intended to improve elementary education, particularly for low achievers. Observations were conducted in 45 third-grade classes using methods similar to those used in the Follow Through study. The ECE program provided for extra aides and greater parent participation in school activities, and the target classes were selected from schools that fell below the 20th percentile in entry level test scores. Thus the students were similar to those in the Follow Through sample, although the ECE classes were taught according to local preference rather than the guidelines of program sponsors.

This study involved both school (not considered here) and class level analyses. The latter were not done on all available variables, but only on a subset of 49 variables selected on the basis of prior research. Of these, 33 showed significant relationships to reading achievement. A few were student-teacher ratio variables indicating that smaller classes generally made greater gains. The rest dealt with classroom activities and teacher-student interaction. Classes that made greater gains spent more time in reading and other academic activities and less in games, group sharing, or socializing. Their teachers spent more time actively instructing in small groups and less time uninvolved with students or involved with individuals rather than groups. They gave more instruction, asked more academic questions, and provided more feedback. Their students asked more questions of their own and initiated more verbal interactions with the teachers.

Clearly, these correlations replicate the Follow Through findings involving student opportunity to learn and active instruction by the teacher. The findings on small class size were not noted in the Follow Through study. Class size has revealed a great range of relationships with achievement in various studies, although meta-analysis suggests that achievement increases as class size decreases (Smith & Glass, 1980). The positive findings for small-group instruction support the first-grade but contradict the third-grade Follow Through data, although the contradiction disappears when the data are interpreted as reflecting the effects of active instruction rather than group size. That is, although instruction can be conducted effectively in either the small-group or the large-group setting, reading achievement gain is linked to frequent active instruction in reading by the teacher.

Another contrast with the Follow Through findings was the absence of significant correlations for level of question (factual vs. open-ended), praise, or criticism. This happened in part because most measures of these variables were not included among the 49 selected for analysis. Also, as noted above, the frequency of academic questions seems to be a more important correlate than either the level of such questions or the nature of the teacher's feedback (praise, acknowledgement, criticism) to the responses that they elicit. In general, then, the Follow Through and ECE studies agree in identifying quantity of academic instruction by the teacher as the key correlate of achievement gain.

Teaching basic skills in secondary schools. Stallings et al. (1978) studied reading instruction at the secondary level, in 27 junior high and 16 senior high reading classes (for low achievers and others who had not yet learned to read efficiently). Instruments were adapted to the activities occurring in these secondary classes, but the same general approach to observation and the same method of observing on three consecutive days were repeated.

Once again, quantity of instruction was the key correlate of achievement. Positive correlates included instructing small or large groups, reviewing or discussing assignments, having the students read aloud, praising their successes, and providing support and corrective feedback when they did not respond correctly. Negative correlates included (1) teacher not interacting with the students; (2) teacher getting organized rather than instructing; (3) teacher offering students choices of activities; (4) students working independently on silent reading or written assignments; (5) time lost to outside intrusions or spent in social interaction; and (6) frequency of negative interactions. In short, gains were minimal when teachers did not concentrate on reading achievement objectives, expected the students to learn mostly on their own, or lost significant instructional time due to disorganization or inability to obtain student cooperation.

Within these general trends, there were differential patterns related to the students' entry-level reading achievement. With students whose functional reading was at a primary level, the most successful teachers tended to use methods traditionally employed in the primary grades, although with more emphasis on comprehension than word attack skills. They would work with one small group while the other students did written work or silent reading. Lessons began with development of vocabulary and concepts, followed by oral reading interspersed with questions to develop and check comprehension. Praise, support, and corrective feedback were frequent. In contrast, teachers working successfully with students who were behind only a grade level or two used methods traditionally employed in the upper grades: less oral reading and more silent reading and written assignments. These teachers still instructed their students actively, however, and structured and monitored their seatwork rather than leaving them mostly on their own.

In summary, across three studies, Stallings and her colleagues found that gains in basic skills achievement were associated positively with active group instruction in the subject matter and negatively with emphasis on nonacademic activities, poor organization or classroom management, or approaches in which students are expected to manage their learning primarily on their own.

Training experiment (secondary reading teachers). Based on the study just described, Stallings developed guidelines for secondary reading instruction (differentiated according to students' entry achievement levels). These guidelines, expressed in terms of percentage of time or frequency per class period, were developed for variables such as instructing individuals, groups, or total class; asking questions; and reacting to students' academic responses and classroom behavior. They provided the basis for an experiment in which the achievement of students of teachers trained to follow the guidelines was compared with that of students in control classes (Stallings, Needels, & Stayrook, 1979).

Analyses indicated that although there was variation in degree of implementation (most of these secondary teachers were not accustomed to having students read aloud, for example, so that this technique was not used as much as it could have been), the treatment teachers eventually approximated the idealized guidelines much more closely than the control teachers did. Furthermore, their students gained an average of six months more in reading achievement (Stallings, 1980). Although not quite statistically significant, this is a sizeable difference and provides some support for the causal efficacy of the behaviors prescribed in the guidelines.

Brophy and Evertson

Brophy, Evertson, and their colleagues completed a series of studies in the 1970s, starting with an assessment of the stability of individual teachers' differential effects on achievement.

Stability study. Brophy (1973) obtained achievement data from students taught during three consecutive years by 88 second-grade and 77 third-grade experienced teachers. Using data from the annually administered Metropolitan Achievement Test (MAT), the students in these 165 teachers' classes were assigned adjusted gain scores on the subtests of word knowledge, word discrimination, reading, arithmetic computation, and arithmetic reasoning (adjustments were based on data for all of the students tested in each year). These adjusted gain scores for individuals then were averaged by class to produce class mean adjusted gain scores for each teacher for each of three consecutive years.

Correlations of these mean adjusted scores from one year to the next (stability coefficients) were low to moderate but positive and usually significant (most were in the .30s). Acland (1976) later reported slightly higher stability coefficients for fifth-grade teachers (averaging .40), and Good and Grouws (1975,1977) reported lower but still statistically significant stability coefficients (averaging .20) for third- and fourth-grade teachers. Thus, investigations of year-to-year stability in teacher effects on student achievement agree in showing that some teachers are consistently better than others at producing student learning gain.

Correlations across the five subsets within each year were considerably higher than the year-to-year stability coefficients for the same subtest. Thus, correlations of word knowledge scores from one year with word knowledge scores from the next tended to be in the .30s, but correlations of word knowledge scores with scores from the other four subtests in the same year were usually much higher, typically in the .70s. Thus, factors unique to a given school year (the teacher's health and welfare, the specific composition and group dynamics of the class, testing conditions, etc.) created cohort effects observable in the achievement data.

Finally, within each class, gains usually were comparable across the two sexes and the five MAT subtests. Few teachers consistently got better results from boys than from girls (or vice versa), or consistently got better results in language arts or reading than in mathematics (or vice versa). These analyses revealed a strong tendency for teachers' effects on achievement to be generalized across the two sexes and the five MAT subtests in any given year, and a weaker but still significant tendency for these general effects to be stable from one year to the next (Brophy, 1973; Veldman & Brophy, 1974). This stability was high enough to allow the next step: process-product research on a subsample of teachers who were unusually consistent in their effects on student achievement.

The Texas Teacher Effectiveness Study. By the time this study was getting organized, achievement data were available for each of the 165 teachers for four consecutive years. Analyses of trends over time indicated that about half of the teachers were stable in their effects on achievement (typically this stability took the form of relative constancy in rank order among the 165 teachers studied, although for a few teachers it took the form of a linear trend indicating steady improvement or deterioration over time). Thirty-one of these consistent teachers were each observed for 10 hours in the first year of this research, and 28 (including 19 holdovers from the first year) were each observed for 30 hours in the second year.

These teachers were selected for *stability* rather than *level* of effectiveness in producing achievement; in fact, as a group they were distributed roughly normally across the range of adjusted MAT means observed in the larger sample of 165. Unfortunately, the district discontinued administration of the MAT prior to the beginning of classroom observation, so that end-of-year achievement data were not available. As a substitute, mean adjusted-gain scores from the four preceding years (for each of the five MAT subtests) were averaged to compute achievement outcome estimates for each teacher. Thus, in this study, process measures were correlated with scores representing predicted effectiveness based on stable prior track records rather than with scores from tests administered subsequent to classroom observations.

Brophy and Evertson relied on an event sampling, in which events relevant to the coding categories are coded when they occur, but nothing is coded when no system-relevant events are occurring. Process data were expressed not only as frequency scores comparable to those used by Flanders and by Stallings, but also as proportion scores (examples: proportion of correct answers followed by praise; proportion of private contacts which dealt with academic work; proportion of these private work contacts which were initiated by the teacher).

Compared to frequency scores, these proportion scores reduce the degree to which measures intended to represent teacher behavior are affected by student behavior. For example, simple frequency scores for teacher praise of good responses are affected by the number of such responses produced. A fast paced class of high achievers might produce 100 correct responses in an hour's lesson; a slower paced group might produce only 40. Frequency scores might reveal that each teacher praises an average of (say) 10 times per hour. These scores will seem to equate the teachers. Proportion scores, however, will reveal that the first praises only about 10% of the students' correct responses, whereas the second praises about 25% (although the frequency data will also be needed to integrate these data fully). Thus, frequency and proportion scores provide different but complementary information.

The presage and process measures generated in this study were analyzed separately for two grade levels (second and third) and two levels of SES to determine relationships to each of the five MAT subtests. The analyses for the two grade levels showed similar patterns of findings and, except for a few measures that were subject-specific in the first place, so did those for the five MAT subtests. However, there were distinctly contrasting patterns of correlates of learning gain for teachers working in low SES versus high SES classrooms. The findings are reported separately, in the form of thousands of correlations (Brophy & Evertson, 1974a; Evertson & Brophy, 1973,1974) and graphs of nonlinear relationships (Brophy & Evertson, 1974b), for the low and high SES subsamples. Brophy and Evertson used the .10 level of significance because of the low sample sizes (18 high SES and 13 low SES classes in the first year, 15 and 13 in the second). However, in interpreting the findings, they stressed general patterns and relationships that held up across both years of the study. Findings that met these criteria are summarized in a book (Brophy & Evertson, 1976).

Presage-outcome data revealed that the teachers who produced the most achievement were businesslike and task oriented. They enjoyed working with students but interacted with them primarily within a teacher-student relationship. They operated their classrooms as learning environments, spending most of their time on academic activities. Teachers who produced the least achievement usually showed either of two contrasting orientations. One was a heavily affective approach in which the teachers were more concerned with personal relationships and affective objectives than with cognitive objectives. The other (fortunately, least common) pattern was seen in disillusioned or bitter teachers who disliked their students and concentrated on authority and discipline in their interviews.

The teachers who produced the most achievement also assumed personal responsibility for doing so. Their interviews revealed (1) feelings of efficacy and internal locus of control; (2) tendency to organize their classrooms and to plan activities proactively on a daily basis; and (3) a "can do" attitude about overcoming problems. Rather than give up and make excuses for failure, these teachers would redouble their efforts, providing slower students with extra attention and more individualized instruction. Such persistence was particularly noticeable among teachers who were successful with low SES students. Here, when there was a poor fit between students' needs and the curriculum's instructional materials and tests, the teachers would often substitute for the materials or develop their own methods of evaluation.

The process variables correlating most strongly and consistently with achievement were those suggesting maximal student engagement in academic activities and minimal time spent in transitions or dealing with procedures or conduct. In general, the successful classroom managers used the techniques described by Kounin (1970) and elaborated by Evertson, Emmer, Anderson, and their colleagues (see Chapter 16 of the *Handbook of Research on Teaching*, in press). They demonstrated "withitness" by monitoring the entire class when they were instructing and by moving around during seatwork time. They rarely made target errors (blaming the wrong student for a disruption) or timing errors (waiting too long to intervene), although they were more likely than other teachers to be coded as overreacting to minor incidents. Even so, they were more likely than other teachers to merely warn rather than threaten their students, and less likely to use personal criticism or punishment. They were proactive in articulating conduct expectations, vigilant in monitoring compliance, and consistent in following through with reminders or demands when necessary.

What these teachers demanded, however, was not so much compliance with authority as productive engagement in academic activities. Such activities were well prepared, and thus ran smoothly with few interruptions and only brief transitions in between. Seatwork assignments were well matched to students' abilities (this typically meant some degree of individualization). Students who needed help could get it from the teacher or some designated person (according to established expectations concerning when and how to seek such help). Students were accountable for careful, complete work, because they knew that the work would be checked and followed up with additional instruction or assignments if necessary. Those who completed their assignments knew what other activity options were available.

There was a difference in emphasis between high SES and low SES classes. The high SES students tended to be eager, compliant, and successful, whereas the low SES students more often were struggling, anxious, or alienated. Consequently, in the high SES classes it was especially important for the teachers to be intellectually stimulating and to provide interesting things for students to do when they finished their assignments, whereas in low SES classrooms it was especially important for the teachers to give students assignments that they could handle and to see that those assignments were done.

Curvilinear relationships were observed between achievement and the percentages of teacher questions that were answered correctly. High SES students progressed optimally when they answered about 70% of these questions correctly, and low SES students when they answered about 80% correctly. These data suggest that learning proceeds most smoothly when material is somewhat new or challenging, yet relatively easy for the students to assimilate to their existing knowledge (even during lessons, when the teacher is present to explain the material and to correct misunderstandings and errors).

Success rates on independent seatwork were not measured, but it was noted that achievement gains were maximized when students consistently completed their work with few interruptions due to confusion or the need for help. This suggested that success rates on these seatwork assignments were high, perhaps approaching 100% (achieved by selecting appropriate tasks in the first place and explaining them thoroughly before releasing the students to work independently). This led the authors to speculate that optimal learning occurs when students move at a brisk pace but in small steps, so that they experience continuous progress and high success rates (averaging perhaps 75% during lessons when the teacher is present and 90-100% when the students must work independently).

Again, there was a relative difference between high and low SES classes: In high SES classes, where most students succeeded with relative ease, the pace could be brisker and the steps slightly larger; in low SES classes, teachers had to move in smaller steps, with more explanation of new material, more practice with feedback, and in general, more redundancy.

Small-group (mostly reading) and whole-class lessons and recitations were common in high gain classes at both SES levels. These lessons often began with presentation of new material or review of old material, and these teacher presentations tended to be rated high in clarity. Then came a practice and feedback phase featuring questions, responses, and feedback. Most questions here were academic, usually low-level or fact questions rather than more open-ended process questions.

In high SES classes, it was important to see that lessons did not become dominated by the most assertive students, by involving everyone, waiting for hesitant students to respond, and insisting that other students refrain from calling out answers. However, it usually was not helpful to question these students repeatedly when they could not answer the original question. Given that most questions were factual and that most of these students were happy to respond if they could, probing in these situations would have amounted to pointless pumping.

Such probing for improved response was effective in low SES classes, however, where many students were anxious or lacking in confidence even when they knew the answers. Here, it was important for teachers to work for any kind of response at all from incommunicative students, and to try to improve the responses of students who spoke up but gave incorrect or incomplete answers. In these situations, giving clues (particularly phonics cues in reading) or rephrasing the question to make it easier were more successful than waiting silently or merely repeating the original question. In contrast to high SES classes, where it was important to suppress unauthorized calling out, called out answers (relevant to the questions asked) correlated positively with achievement in low SES classes.

Surprisingly, the use of patterned turns in small groups (mostly reading groups) correlated positively with achievement. That is, teachers who went around the small group in order, giving each successive student a turn, got greater gains than teachers who randomly called on students or called primarily on volunteers. One probable reason for this is that the patterned turns mechanism insured that all students participated regularly and roughly equally. Furthermore, in high SES classes, it helped focus students' attention on the content of the lesson rather than on attempts to get the teacher to call on them, and in low SES classes, it provided structure and predictability that may have been helpful to anxious students.

The correlations involving motivation variables were generally much weaker than those involving classroom management and academic instruction variables. Positive correlations were obtained in both SES levels for use of symbolic rewards, especially stars or smiling faces on papers that could be taken home to show parents. Concrete rewards or tokens were not used in any systematic way by the teachers under study. The findings for academic praise and criticism varied by SES and by teacher versus student initiation of interaction. Praise given in teacher initiated interactions was widely distributed and correlated positively with achievement. However, praise given during student initiated interactions went mostly to those students who frequently approached the teacher to show their work, and such praise correlated negatively with achievement. In general, measures of academic praise correlated positively but weakly in low SES classes, but were unrelated to or negatively (and again, weakly) correlated with achievement in high SES classes.

Criticism for poor academic responses or poor work correlated positively with such gain (in high SES classes only). As in the Stallings work described above, such academic criticism was rare, so that the correlation is based on the difference between rarely criticizing students for working below their abilities and never doing so.

Academic praise was much more frequent than academic criticism, but this was not true for teachers' responses to student conduct. In fact, praise of good conduct was very rare and never correlated significantly with achievement. Criticism and punishment for misconduct were more frequent, however, and tended to correlate negatively with achievement. The teachers who elicited greater achievement tended to respond to misconduct with simple directives or warnings rather than with criticism or punishment. When something more was required, they tended to arrange an individual conference to discuss the problem and come to some agreement with the student about what was to be done. They were unlikely to lash out at students, to punish them impulsively, or to send them to the principal for discipline.

In general, the teachers who got the most gain in high SES classes motivated students by challenging and communicating high expectations to them, occasionally delivering symbolic rewards when the students succeeded and, on rare occasions, criticizing them when they failed due to inattentiveness or poor effort. In contrast, the teachers who got the most gains in low SES classes motivated students primarily through gentle and positive encouragement rather than challenge or demand. They not only used symbolic rewards, but often praised their students within the contexts of personalized interactions with them.

The following variables failed to correlate significantly with achievement: teachers' warmth and enthusiasm; components of Flanders' indirectness (use of student ideas, frequent student-student interaction); advance organizers; ratio of divergent to convergent questions; democratic leadership style; confidence; and politeness to students. Brophy and Evertson (1976) argued that variables such as warmth and politeness should be expected to relate more to attitudes than achievement. For other variables (enthusiasm, advance organizers, indirectness), they argued that significant correlations did not appear because the data had been collected in the primary grades, where (1) students tend to be positively oriented toward and accepting of teachers and the curriculum (so that enthusiasm is not of great importance) (2) presentations tend to be short and concentrated on isolated facts (so that advance organizers are less important), and (3) instruction focuses on basic skills rather than use of these skills to deal with more abstract and intellectual content (so that instruction and supervision of practice is more important than teacher use of student ideas or stimulation of student-student discussion). In short, they argued, some of the classroom processes that are frequent and important for learning in the primary grades are infrequent and unimportant in other grades, and vice versa.

Junior high study. These speculations about grade level differences were tested in a follow up study at the junior high level (seventh and eighth grade), using methods similar to those used in the second- and third-grade study but adapted to include measures of time spent in various activities (Evertson, Anderson, & Brophy, 1978; Evertson, Anderson, Anderson, & Brophy, 1980; Evertson, Emmer, & Brophy, 1980). Thirty-nine English and 29 mathematics teachers were observed an average of 20 times in each of two class sections (total N = 136 classes). These included most of the English and mathematics teachers working in nine of the city's 11 junior high schools (the other two, which happened to be the lowest in average SES level, were excluded because they used individualized mathematics programs that could not be studied with the same methods).

Entry level achievement was measured by the English and mathematics subtests of the California Achievement Test (CAT) given the previous spring. Achievement during the observation year was measured with specially prepared tests based on the content actually taught in these classes. The CAT scores accounted for 71% of the variance in end-of-year achievement in mathematics, and 85% in English. Students were also asked to rate how likeable and accessible the teachers were, how much they profited from the class, how likely they were to choose this teacher again, and so on. Factor analysis of these nine ratings produced a strong first factor, which was used as a measure of student attitude. These attitude scores correlated positively (.32) with adjusted achievement in mathematics but negatively (-.24) in English.

Because data were available on two class sections for each teacher, it was possible to compute correlations reflecting stability of teacher effects across classes within the same year. In mathematics, these correlations were .37 for adjusted achievement and .44 for attitude. When the data for five teachers whose two mathematics sections differed by more than 40 points on the CAT (approximately two grade equivalents) were removed, these correlations rose to .57 for achievement and .57 for attitudes. Thus, the stability of teacher effects on junior high mathematics achievement across class sections within the same year was higher than the stability across successive years observed earlier in the second- and third-grade study, and stability of effects on attitude was even higher. Also, attitude was correlated positively with achievement.

The data for the English classes were more complex. Here, stability correlations were only .05 for achievement but .82 for attitude. These rose to .29 and .83, respectively, when data from the 13 English teachers with highly contrasting class sections were removed (Emmer, Evertson, & Brophy, 1979). Thus, effects on achievement were not stable and were correlated negatively with effects on attitudes (attitude effects were highly stable, however). Given that 85% of the variance in adjusted achievement in English was accounted for by CAT scores, there was little reliable variance left to be explained by classroom process measures. The root problem here was that a great range of academic content and activities appeared in these classes, despite their ostensible comparability. Some teachers concentrated on grammar and basic skills, others on reading comprehension or composition, and still others on poetry or drama. This range of activities minimized the degree to which the end-of-year tests could sample from a rich pool of common learning objectives. Thus, despite efforts to avoid this problem by monitoring the content taught, it was not possible to devise a test that would be both valid and discriminating for evaluating achievement in these English classes.

Only two general process-product patterns emerged in English classes: achievement was greater where serious misbehaviors were uncommon and where teacher praise during class discussions was relatively frequent. There also were some findings that applied only to the classes that were below average in CAT scores. Greater gains were made in these lower ability classes when the teachers (1) were friendlier and more accepting of students' social initiations and personal requests; (2) encouraged students to express themselves, even to the extent of tolerating relatively high rates of calling out; and (3) were, nevertheless, relatively strict disciplinarians. As far as they go, these data from low ability junior high English classes are similar to the data from low SES second and third grade classes.

Students in English classes expressed positive attitudes toward teachers who were rated (by observers) as warm, nurturant, enthusiastic, and oriented to students' personal needs who provided more choice and variety in assignments. The students had less positive attitudes toward teachers who were academically demanding, used extensive discussion, asked difficult questions, or criticized or tried to improve unsatisfactory responses. In general, English classes in which the teacher was perceived as "nice" and the class as enjoyable but undemanding produced the most positive attitudes.

In mathematics, there was much more overlap between the processes associated with achievement and those associated with positive attitudes. Classroom organization and instruction variables correlated more strongly with achievement, and measures of teachers' personal qualities correlated more highly with student attitudes, but, in general, the correlations were in the same direction. The more popular mathematics teachers not only had good relationships with their students but were academically stimulating and demanding.

The more successful mathematics teachers were rated highly as classroom managers, even though behavior problems were observed just as often in their classes as in others. Perhaps they were better at "nipping problems in the bud" by stopping them quickly before they go out of hand. In any case, variables like monitoring (withitness) and avoidance of target and timing errors were important, especially in the low ability classes.

Measures of the amount and quality of instruction were even more directly related to achievement in these classes than they were in the second- and third-grade classes studied earlier. The more successful teachers taught more actively, spending more time lecturing, demonstrating, or leading recitation or discussion lessons. They devoted less time to seatwork, but were more instructionally active during the seatwork time they did have, being more likely to monitor and assist the students rather than leave them to work without supervision.

Concerning teacher questioning, the major difference was quantitative: The more successful teachers asked many more questions. Most of these were product rather than process questions, although in contrast to the findings from the early grades, the percentage of total questions asked that were process questions correlated positively with achievement in these junior high mathematics classes. About 24 questions were asked per 50-minute period in the high gain classes, and 25% of these were process questions. In contrast, only about 8.5 questions were asked per period in the low gain classes, and only about 15% of these were process questions.

There were no clear findings for difficulty level of question (as represented by the percentage of questions answered correctly rather than by the distribution of type of question; process questions are not necessarily harder than product questions). However, student failure to make any response at all (in contrast to responding substantively but incorrectly) was negatively correlated with achievement, again indicating the importance of teachers' getting some kind of response to each question asked.

Small-group instruction was virtually absent from these classes, so that the "patterned turns" variable was irrelevant. Most lessons were with the whole class, and response opportunities were usually created by calling on nonvolunteers (45%), calling on volunteers (25%), or accepting call-outs (25%). Of these, calling on volunteers correlated positively with achievement. Calling on nonvolunteers was not particularly harmful, at least when they were following the lesson and likely to know the answer. However, high rates of calling on nonvolunteers who then answered incorrectly were associated negatively with achievement. Similarly, call-outs were not particularly harmful so long as the teacher retained control over participation in the lesson. High call-out rates suggested absence of such control, but many teachers with intermediate rates used call-outs effectively to keep the class moving or to encourage student participation (especially in low ability classes). Accepting called out questions or comments was associated positively with achievement in the low ability classes.

Public praise of good answers was low key and infrequent, but it correlated positively (although weakly) with achievement. Praise during private interactions, criticism of poor answers or poor work, and attempts to improve unsatisfactory responses were all unrelated to achievement. In general, unlike the primary grades where it is essential to take the time to work with individuals during (small-group) lessons, in the upper grades it is more important to keep (whole-class) lessons moving at a brisk pace.

Use of students' ideas (redirection of their questions to the class and integration of their comments into the discussion) related positively. Thus, except for student-student interaction, key elements of Flanders' concept of indirectness (teacher questions, praise, and use of student ideas) were associated positively with both achievement and attitude in this study. Note, however, that these events occurred within the context of teacher-directed, whole-class instruction on academic content. Furthermore, other positive relationships were observed for emphasis on active instruction (lecture-demonstrations, time spent in the developmental portion of the mathematics lesson). Thus, aspects of what Flanders called "indirect" instruction complement and co-occur with aspects of what others have called "direct" instruction. Both are aspects of what Good (1979) has called "active" instruction, and they contrast not so much with each other as with patterns in which the teacher does not instruct at all or expects the students to learn primarily on their own.

The more successful teachers had more frequent but shorter individualized contacts with students during seatwork times. This probably was because they did not release their students to begin the work until it had been explained thoroughly, so the students needed less reteaching later. Also, these teachers were generally "withit," and one aspect of this is keeping track of the whole class rather than becoming too involved for too long with individuals.

Correlations involving high inference ratings indicated that the observers saw these successful mathematics classes as follows: Teacher maintains order and commands respect; teacher monitors class and enforces rules consistently; transitions are efficient and disruption infrequent; and teacher appears competent, confident, credible, enthusiastic, receptive to student input, and clear in presentations. Successful teachers were also rated higher on items dealing with expectations and academic orientation: academic encouragement, concern for achievement and grades, well prepared, uses available time for academic activities.

Taken together, the data from this study suggest resolutions to certain apparent discrepancies in previous findings. Along with Stallings' data on secondary remedial reading classes, these data from junior high mathematics classes show that linkages between achievement and measures of opportunity to learn, efficient classroom management, and active instruction by the teacher apply to the late elementary and secondary grades as well as to the primary grades and to classes in all kinds of schools, not just those serving low SES populations. On the other hand, the limited findings for the English classes remind us that these linkages do not appear for certain learning objectives or when there is poor overlap between what is taught and what is tested. They appear most clearly in studies where the objectives involve knowledge and skills that can be taught specifically and tested by requiring students to reproduce them.

The junior high mathematics data also show how classroom processes and process-product relationships vary with grade level. The primary grades stress instruction in basic skills, and it is important to see that each student participates actively in lessons and gets opportunities to practice and receive feedback. In the higher grades, more time is spent learning subject matter content, and students are more able to learn efficiently from listening to the teachers' presentations or to exchanges between the teacher and other students. There is less need for small group instruction and for overt involvement of each student. However, it is important that teachers maintain attention to well prepared and well paced presentations, and that these presentations be clear and complete enough to enable the students to master key concepts and apply them in follow up assignments. These grade level differences account for most of the apparent discrepancies in process-product findings. Few such findings are contradictory, but most need qualification by grade level and other context factors.

First-grade reading group study. Brophy and Evertson and their colleagues also completed an experimental study of first grade reading instruction (Anderson, Evertson, & Brophy, 1979), using a small-group instruction model based on their own process-product work and on early childhood education programs developed by Blank (1973) and by the Southwest Educational Development Laboratory (1973). The model was not specific to reading instruction; instead, it was intended for any small-group instruction that called for frequent recitation or performance by students. It consisted of 22 principles for organizing, managing, and instructing the group as a whole, and for providing feedback to individual students' answers to questions. These principles, along with brief explanations, were organized into a manual that provided the basis for the treatment. In October, each treatment-group teacher met with a researcher who described the study and presented the manual. The researcher returned a week later to administer a test of the teacher's mastery of the principles, and to discuss any questions or concerns.

Classes from nine schools serving predominantly middle class Anglo populations were assigned randomly (by school) to one of three groups (all classes in any given school were in the same group). Treatment-observed (N = 10) classes received the treatment and were observed periodically throughout the year. Treatment-unobserved classes (N = 7) received the treatment but were not observed. Control classes (N = 10) did not receive the treatment but were observed. Inclusion of the treatment-unobserved group allowed for assessment of the possible effects of observer presence on treatment effects, and inclusion of classroom observation in both treatment and control classes allowed for assessment of treatment implementation and process-product relationships in addition to effects on achievement (adjusted for entry level reading readiness).

From November and through April, the 10 treatment-observed classes and 10 control classes were observed about once a week, with emphasis on behaviors relevant to the principles in the model. These principles concerned managing the group efficiently, maintaining everyone's involvement, and providing for sufficient instruction, practice, and feedback for each individual within the group context. The teachers were advised to: sit so that they could monitor the rest of the class while teaching the reading group, begin transitions with a standard signal and lessons with an overview of objectives and a presentation of new words, prepare the students for new lesson segments and seatwork assignments, call on each individual student for overt practice of any concept or skill considered crucial, avoid choral responses, apportion reading turns and response opportunities by the patterned-turns method rather than by calling on volunteers, discourage call-outs, wait for answers, and try to improve unsatisfactory answers when questions lent themselves to rephrasing or giving of clues.

Praise of good performance was to be used only in moderation and was to be as specific and individualized as possible. Academic criticism (not mere negative feedback) was to be minimized but, if given, was to include specification of desirable or correct alternatives. If the students were progressing nicely through the lesson as a group, they were to be kept together. If not, the teacher was to dismiss those who had mastered the material and work more intensively with those who needed extra help.

Achievement data indicated that both treatment groups outperformed the control group, and that these treatment effects did not interact with entering readiness levels (class averages). There was no difference between the two treatment groups, indicating that the presence of classroom observers did not affect the results and was not necessary for treatment effectiveness.

The treatment was implemented unevenly. The best implemented principles were those calling for frequent individualized opportunities for practice, minimal choral responses, use of ordered turns, frequent sustaining feedback, and moderate use of praise. In general, these well implemented principles also correlated as expected with achievement. Not well implemented were the suggestions about beginning with an overview, repeating new words, giving clear explanations, and breaking up the group. With hindsight, some of these guidelines seem unnecessary or irrelevant to first-grade reading group instruction, and others seem unlikely to be implemented without a more powerful treatment.

Process-product data revealed greater achievement gains where more time was spent in reading groups and in active instruction, and less time was spent dealing with misbehavior; transitions were shorter; the teacher sat so as to be able to monitor the class while teaching the small group; lessons were introduced with overviews; new words were presented with attention to relevant phonics cues; lessons included frequent opportunities for individuals to read and to answer questions about the reading; most questions called for response from an individual rather than from the group; most responses resulted from ordered turns rather than volunteering or calling out; most incorrect answers were followed by attempts to improve the response through rephrasing the question or giving clues; occasional incorrect answers were followed by detailed process explanations (in effect, reteaching the point at issue); correct answers were followed by new questions about 20% of the time rather than less frequently; and praise of correct responses was infrequent but relatively more specific (although the absolute levels of specificity of praise were remarkably low, even for the treatment teachers). Group call-outs were associated positively with achievement for the lower ability groups and negatively for the higher ability groups. Anderson, Evertson, and Brophy (1982) have revised and reorganized their guidelines for first-grade reading group instruction based on these findings from this study. These guidelines summarize the apparent implications of the findings for practice (see Anderson, Evertson, & Brophy, 1979, for detailed presentation of the findings themselves and see Appendix for principles).

Good and Grouws

Good and Grouws and their colleagues also conducted process-outcome research in different settings and then developed and tested a teaching model (in this case, for whole-class instruction in mathematics).

Stability analyses. The work began with collection of attitude and achievement data for two consecutive years for most of the third- and fourth-grade teachers (N = 103) in a predominantly white, suburban school district. Year-to-year stability coefficients for adjusted achievement gain on subtests of the Iowa Tests of Basic Skills were statistically significant but low, averaging only about .20 (Good & Grouws, 1975). These teachers did a great deal of formal and informal sharing of students, which may explain why the stability coefficients were lower than those typically obtained from classrooms in which the teachers work with the same students all day in all subjects. Stability coefficients for classroom climate (attitudes toward the teacher and the class) were also low (averaging .22), perhaps because attitudes were generally quite positive (so the variance was restricted).

Achievement and attitude measures were uncorrelated. Consequently, the original plan to select teachers who were stable in their effects both on attitudes and on achievement in various subject matter areas had to be abandoned in favor of concentration on a single subject. Good and Grouws selected mathematics, partly because stability coefficients were somewhat higher in this subject. They identified nine fourth-grade teachers who taught mathematics to the same students throughout the year and whose classes were in the top third in adjusted achievement in both years and nine parallel teachers whose classes were in the lower third in both years. These 18 teachers (and, in fact, all fourth-grade teachers in the district) used the same textbook.

Fourth-grade naturalistic study. The following fall, these 18 teachers were each observed seven times. Mathematics achievement on the Iowa Tests of Basic Skills was measured in the fall and again in the spring. In addition, to protect the anonymity of the 18 selected teachers, the same process and product data were collected in an additional 23 fourth-grade classes. Thus, the data include correlations for the total sample of 41 classes, as well as comparisons of the nine high scoring teachers' classes with the nine low scoring teachers' classes. The correlational data will be discussed in a later section in conjunction with data from subsequent research in low SES classes. For now, consider the data from the 18 selected teachers. These teachers maintained their relative positions in the third year: Once again, teachers of the nine high scoring classes elicited considerably greater achievement gain from students than teachers of the nine low scoring classes.

All 18 teachers used whole-class instruction followed by seatwork/homework assignments (the teachers who subdivided their classes into groups for differentiated instruction and assignments tended to elicit medium levels of achievement gain, as did some teachers who used the whole-class method). Thus, neither the whole-class nor the small-group method was clearly superior. Teachers who got the best results used the whole-class method, but so did teachers who got the worst results. Good and Grouws (1975,1977) argue that the whole-class method is more efficient for fourth-grade mathematics instruction when used effectively, but note that it requires classroom management and instruction skills that many teachers do not possess.

Teachers who elicited higher achievement from their students had better managed classes even though they had more students. They spent less time in transitions and disciplinary activity, and their students called out more answers, asked more questions, and initiated more private academic contacts with the teachers. Classroom climate ratings and student attitudes were more positive in these classes, even though the teachers' emphasis was clearly on academics.

Teachers of higher achieving classes moved through the curriculum at a brisker pace. They covered an average of 1.13 pages per day, compared to only 0.71 for teachers with lower achievement gain classes (Good, Grouws, & Beckerman, 1978). Page coverage correlated .49 with achievement.

Teachers of higher achieving classes instructed more clearly and introduced more new concepts in the development portions of lessons. The pace was quicker, and less time was spent going over previous assignments. In contrast, teachers of lower achieving classes provided less clear instruction, so that, by inference, more of their instructional attempts came in the form of corrections of unsatisfactory responses to questions or assignments.

Teachers of the high achievement gain classes asked fewer questions (probably because they spent less time going over mistakes made on previous assignments). In particular, they asked fewer questions that yielded incorrect answers or failures to respond. When errors or response failures did occur, however, these teachers were twice as likely to give process feedback (explain the steps involved in developing the answer) as they were to merely supply the correct answer. Their lessons moved at a brisker pace, then, for several reasons. First, they made clearer presentations at the beginning. Second, they "kept the ball moving" by interweaving explanations with questions, rather than relying more heavily on recitation. Third, more of their questions were direct, factual questions likely to produce immediate correct answers. Fourth, when students were confused, these teachers would revert to explanation rather than merely providing correct answers or attempting to elicit them through continued questioning.

During seatwork times, teachers of higher achieving classes circulated to monitor progress. Yet, they averaged only three teacher-initiated work contacts (but 23 student-initiated work contacts) per hour, compared to averages of 6 and 12, respectively, for teachers of the low achieving classes. Thus, they concentrated on giving help where it was most needed. Furthermore, their feedback during these private contacts was more likely to involve explanation (not just giving the answer or brief directives).

Good and Grouws (1977) describe the feedback of teachers of high achieving classes as immediate, nonevaluative, and task-relevant. These teachers both praised and criticized less than teachers of low achieving classes, and their evaluative responses were more contingent on quality of performance (teachers of the lower achieving classes frequently praised students for something other than correct performance).

Summarizing their findings, Good and Grouws (1977) state that the higher achieving classes showed the following clusters: frequent student initiation of academic interaction; whole-class instruction; clarity of instruction, with availability of information as needed (process feedback in particular); nonevaluative and relaxed, yet task-focused learning environments; higher achievement expectations (faster pace, more homework); and relative freedom from disruption. Even so, the effectiveness of these teachers was not always immediately obvious. Naive observers regularly rated teachers of the lower achieving classes as low, but rated many of the teachers of higher achieving classes as average rather than high. Thus, although low teacher effectiveness is easy to spot because of poor management or lack of much instruction at all, observers may need training in what to look for in order to identify teachers who maximize student achievement gain.

Fourth-grade experimental study. Good and Grouws (1979b) next conducted a treatment study, still in fourth-grade mathematics but this time in urban schools serving primarily low SES families. The treatment involved a set of instructional principles organized into a model (shown in summary form in Table 3) calling for briskly paced whole-class instruction supplemented by homework assignments.

The model prescribes more active whole-class instruction than most teachers deliver (particularly in development portions of lessons) and more frequent reviewing. Less time is allocated for going over homework and less time is spent on seatwork. The emphasis on development and review and the inclusion of mental computation exercises were based on previous mathematics education research suggesting that many teachers rely too much on independent seatwork (often without sufficient monitoring, accountability, or follow up), and that students need more extensive development of concepts, better advance structuring and subsequent follow up of assignments, and more opportunities to think about and integrate mathematical concepts. Consequently, these elements were added to the model and integrated with elements drawn from the previous process-product study (whole-class approach, brisk pacing, programming for high success rates, active instruction, homework assignments).

Manuals explaining the model were given to the 21 treatment teachers and were discussed in two 90-minute meetings. The investigators also met with the 19 control teachers, not to give specific guidelines about instruction, but to explain the importance of the study and to heighten their attention to and

Table 3

Good and Grouws' (1979) Guidelines for
Fourth-Grade Mathematics Instruction

Summary of Key Instructional Behaviors

Daily Review (First 8 minutes except Mondays)

a. review the concepts and skills associated with the homework
b. collect and deal with homework assignments
c. ask several mental computation exercises

Development (About 20 minutes)

a. briefly focus on prerequisite skills and concepts
b. focus on meaning and promoting student understanding by using lively explanations, demonstrations, process explanations, illustrations, etc.
c. assess student comprehension
    1. using process/product questions (active interaction)
    2. using controlled practice
d. repeat and elaborate on the meaning portion as necessary

Seatwork (About 15 minutes)

a. provide uninterrupted successful practice
b. momentum--keep the ball rolling--get everyone involved, then sustain involvement
c. alerting--let students know their work will be checked at the end of the period
d. accountability--check the students' work

Homework Assignment

a. assign on a regular basis at the end of each math class except Fridays
b. should involve about 15 minutes of work to be done at home
c. should include one or two review problems

Special Reviews

a. weekly review/maintenance
    1. conduct during the first 20 minutes each Monday
    2. focus on skills and concepts covered during the previous week
b. monthly review/maintenance
    1. conduct every fourth Monday
    2. focus on skills and concepts covered since last monthly review

enthusiasm about their mathematics instruction. This was intended to minimize the degree to which outcomes favoring the treatment group could be attributed to Hawthorne effects associated with participating in an experiment.

From October through late January, each treatment and control teacher was observed six times. Most (19 of 20) treatment teachers implemented most program elements. The major exception was development, which usually was no more extensive in the treatment than in the control classes. The treatment classes outperformed the control classes both on a standardized mathematics test (SRA, Short-Form E, Blue Level) and on a criterion-referenced test of the content actually taught during the observation period. Student attitude data also favored the treatment classes.

Achievement gains were substantial. In a few months, the treatment group increased from the 27th to the 58th percentile on national norms, and the teachers who had the highest implementation scores produced the best results. The control group's performance did not match that of the treatment group, but it exceeded expectations based on previous years. This improvement may have been due to Hawthorne effects associated with the authors' attempt to develop heightened enthusiasm about mathematics instruction. Interviews revealed that the control teachers had not been exposed to the treatment nor changed their previous teaching behavior in major ways, but that they had thought more about their mathematics instruction. Of these 19 control teachers, 12 used the whole-class approach and 7 used small groups.

Subsequent analyses (Ebmeier & Good, 1979) indicated that main effects on achievement were elaborated by interactions with teacher (four types) and student (four types) characteristics. For example, the performance of low achieving and dependent students (especially when taught by certain types of teachers) was particularly enhanced by the treatment relative to that of higher achieving and independent students. Also, teachers classified as "unsure" benefited more than those classified as "secure." Thus, the treatment was especially effective with both teachers and students who needed more structure.

Other treatment studies. Good and Grouws completed two more treatment studies at Grade 6 (Good & Grouws, 1979a), and at Grades 8 and 9 (Good & Grouws, 1981). In these studies, the treatment included not only the model shown in Table 3, but also a supplementary model for teaching verbal problem solving. These studies are not described in detail here because they are highly specific to mathematics instruction (see Chapter 35 of the *Handbook of Research on Teaching)*. In general, their effects were positive but weaker than those seen in the fourth-grade treatment study, mostly because treatment implementation was less consistent. This work on what has been called the Missouri Math Program is summarized in *Active Mathematics Teaching* (Good, Grouws, & Ebmeier, 1983).

High SES versus low SES comparisons. Good, Ebmeier, and Beckerman (1978) presented data from the fourth-grade naturalistic study (Good & Grouws, 1977) and treatment study (Good & Grouws, 1979b) that allow comparisons with the SES difference findings reported by Brophy and Evertson (1974b, 1976), although each data set has unique aspects. The teachers in Good and Grouws's naturalistic study include the nine consistently high achieving and nine consistently low achieving teachers who used the whole-class approach, plus other teachers who were less consistent and extreme in their effects on achievement (many of whom used the small-group approach). They all taught in suburban schools. The 40 teachers in the experimental study included 21 who were implementing the treatment model and thus behaving differently than they would have otherwise. They taught in an urban district. The Brophy and Evertson data, in contrast, included teaching in all subject areas (not just mathematics) in second and third grade in an urban district. The teachers were stable in their effects on achievement, but distributed normally in degree of effectiveness.

Good, Ebmeier, and Beckerman (1978) note that the process-outcome correlations in their studies are generally lower than those involving similar variables from the Brophy and Evertson study. One possible reason is lower reliability of the process measures. The teachers in the two studies described by Good, Ebmeier and Beckerman were observed for less time and only during mathematics. Therefore, some behaviors may not have occurred often enough to allow reliable measurement. Also, all of the teachers in the Brophy and Evertson study had demonstrated stability in effects on achievement and may also have been unusually stable in their classroom behavior. This was true for only 18 of the teachers studied by Good and Grouws. Also, both fourth-grade mathematics samples contained a majority of teachers who taught the whole class and a minority who used small groups. It is likely that ostensibly identical classroom process measures actually had different meanings and patterns of correlation with outcomes in these two types of classes.

As an example, consider the data on development portions of lessons. In the naturalistic study, teachers of the nine higher achieving classes spent somewhat more time in development than teachers of the nine low achieving classes did, yet the correlation between development time and achievement for the sample as a whole was -.13. Similarly, although the guidelines for development time were poorly implemented in the treatment study, the correlation between development time and achievement time here was -.14. Two factors contributed to these anomalous findings. First, the measure of development was quantitative (time). There is no necessary relationship between time spent in development and the quality of that development (clarity, completeness, focus on the right concepts at the right level of detail). Second, the teachers who used small groups were among those with the highest development time, because they taught several small group lessons that each included some introductory lecture or presentation. Much of this was redundant with what was said in their other small-group lessons, but it nevertheless counted as development time. Problems of this sort may have existed with other process measures as well.

Besides showing fewer significant relationships, these fourth-grade mathematics data differed from Brophy and Evertson's data in that most relationships held up across the two SES settings. The SES differences that did appear, however, were generally similar to those reported by Brophy and Evertson. Both sets of data indicate that it was essential for teachers in low SES classes to regularly monitor activity, supervise seatwork, and initiate interactions with students who needed help or supervision. Teachers in high SES classes did not have to be quite so vigilant or initiatory and for the most part could confine themselves to responding to students who indicated a need for help. Positive affect, a relaxed learning climate, and praise of student responses were also more related to student achievement in low SES settings. An academic focus, which included frequent lessons involving questioning the students, was associated with achievement in both settings, although in low SES settings it was important that most questions be factual, product questions rather than more open-ended process questions. Similar findings were reported by Soar and Soar (1979).

The only clear contradiction noted by Good, Ebmeier, and Beckerman (1978) involved a set of (mostly nonsignificant) trends indicating that it was more often advisable to try to improve unsatisfactory responses to questions in the high SES than in the low SES classes. Brophy and Evertson found the opposite and suggested that, given the factual nature of most questions in the early grades and the eagerness of most high SES students to respond, most teacher attempts to improve student failure to respond would amount to pointless pumping. It is possible that by fourth grade, and especially in mathematics (a subject that is difficult for many students and lends itself well to rephrasing of questions or provision of clues), it is the bright and eager students who profit most from attempts to improve responses and the slowest and most anxious students for whom such attempts would be pointless pumping. In any case, issues concerning when and how teachers should try to improve responses seem unlikely to be resolved until they are attacked with qualitative rather than just quantitative measures.

Beginning Teacher Evaluation Study (BTES)

In 1970, the state of California established a commission to oversee teacher education and certification programs in the state. In 1972, the commission began planning a study to identify teaching competencies that could be used as the basis for evaluating beginning teachers. As planning progressed, however, discussion began to focus more on the need for research linking teacher behavior to student achievement. Eventually, with funding from the National Institute of Education and participation by researchers from the Educational Testing Service and the Far West Regional Laboratory for Educational Research and Development, a series of studies was conducted (Powell, 1980). Although the BTES name was applied to this series collectively, the studies involved experienced rather than beginning teachers and concentrated on research rather than evaluation.

BTES Phase II: first field study. During 1973-1974, data were collected in 41 second-grade and 54 fifth-grade classes. The teachers had at least three years of experience and worked in a variety of school districts. Data were collected on teachers' aptitudes, diagnostic skills, knowledge about subject matter, expectations, preparation for instruction, and behavior, and on students' aptitudes, cognitive styles, expectations, and achievement. Classes were observed using two low inference systems. One (the "RAMOS" system) focused on the teacher and the nature of the instruction occurring at the time, and the other (the "APPLE" system) focused on the activity of eight target students stratified by sex and achievement level. The RAMOS system was used during reading and mathematics instruction, and the APPLE system throughout the school day. Most teachers were observed four times, twice with each system. The data are presented in a five-volume final report (McDonald & Elias, 1976b), in a summary report (McDonald & Elias, 1976a), and in briefer publications (McDonald, 1976,1977).

The findings are difficult to summarize and compare with data from related studies for several reasons. First, although sophisticated statistical methods (including multiple regression and path analysis) were used, the reports do not include correlations or other statistics linking each separate process variable to achievement. Instead, each analysis gives information about only a few process variables--those that added significantly to the variance in achievement accounted for by multiple correlations (i.e., those whose partial correlations with adjusted achievement remained significant when the effects of all other predictors were controlled). Second, although it picked up dyadic teacher-student interaction data comparable in some ways to the data developed in the Brophy and Evertson and the Good and Grouws studies, the APPLE system placed the student in the foreground. Detailed information about the teacher's behavior appeared only when the teacher happened to be interacting with a target student when that student was being observed. Third, most of the process variables used in the analyses were combination scores that lumped together different teacher behaviors (for example, time spent disciplining or preparing to instruct was aggregated with time spent actually instructing in a measure of "direct teaching time"). Consequently, the data from Phase II of BTES cannot be compared directly with the work reviewed so far.

Still, certain general trends are familiar. The largest adjusted achievement gains occurred in classes of teachers who were well organized, who maximized the time devoted to instruction and minimized time devoted to preparation, procedure, and discipline, and who spent most of their time actively instructing the students and monitoring their seatwork. Their students were mostly attentive to lessons and engaged in their assignments when working alone. Time spent overtly practicing specific skills (such as word attack in reading or computation in mathematics) was positively correlated with achievement in second grade. By fifth grade, time spent in these basic skills was negatively associated with achievement, but time spent in lessons on applications of these skills (reading comprehension, mathematics problem solving) was positively associated. Positive feedback and praise were positive correlates in second-grade reading and fifth-grade math. Variety of materials was a positive correlate in second-grade reading but a negative correlate in the other three data sets.

Even though general trends could be identified, none of the teacher behavior measures was a significant predictor of achievement for both subject matters (reading, mathematics) at both grade levels (second, fifth). Thus, the data did not support a basic assumption that had led to the BTES in the first place, the notion that there are generic teaching skills that are appropriate and desirable in any teaching situation. Most other data also support this conclusion. Although certain abstract principles appear to be universal (e.g., match difficulty level of content to students' present achievement levels), few if any specific, concrete teacher behaviors are generic correlates of achievement (see Gage, 1979, on this point).

BTES Phase III-A: ethnographic study. During 1974-1975, Phase III-A of BTES included ethnographic study of the classes of 20 second-grade and 20 fifth-grade teachers in the BTES "known sample." This sample had been culled from larger samples of 100 teachers at each grade level based on data from special two-week units in reading and mathematics. The 40 teachers in the "known sample" consisted of 10 at each grade level considered to be "more effective" and 10 considered "less effective" on the basis of teacher behavior and student achievement in these special units.

Unlike most research reviewed here in which data gathering was focused on previously specified events (usually, ongoing events were coded into categories in low inference coding systems), this study used the thick description, "ethnographic" method in which observers record free form, running descriptions of events as they occur (see Chapter 5 of the *Handbook for Research on Teaching*). Heretofore, ethnographic methods have been used mostly in case studies of just one or a small number of classes. In Phase III-A of BTES, however, these methods were used in large enough samples of comparable classrooms to allow the use of inferential statistics.

This process was as follows. First, ethnographers (mostly graduate students in sociology and anthropology) were recruited, familiarized with second- and fifth-grade classrooms, and trained to write protocols describing reading and mathematics instruction. Then, the ethnographers visited the classes for a week at a time, typically observing two more effective and two less effective teachers at the same grade level (the ethnographers were not told how the teachers had been classified). Notes from these observations were then tape recorded and transcribed, and raters representing different types of expertise studied pairs of protocols (one from a more effective teacher and one from a less effective teacher) and generated dimensions on which the larger set of protocols might be compared. Eventually, 61 such dimensions were identified and rated in each protocol.

The final data were generated by training new raters to consider pairs of protocols (again, one of each pair was from a more effective teacher and one from a less effective teacher, but raters did not know which was which) and determine which protocol gave more evidence of the behavior described by each of the 61 variables. There were 100 pairings possible at each grade level (each of 10 more effective teachers could be paired with each of 10 less effective teachers). Of these, randomly selected samples of 36 pairings were rated for each subject matter at each grade level. The data are presented in a technical report (Tikunoff, Berliner, & Rist, 1975) and in subsequent publications (Berliner & Tikunoff, 1976,1977).

In contrast to the BTES Phase II data (on teachers who were not selected on the basis of previously demonstrated effectiveness), these data on the BTES "known sample" yielded many findings that held up across both grade level and subject matter. Twenty-one of the 61 variables yielded significant differences in all four data subsets (second-grade reading, fifth-grade reading, etc.). All 61 variables showed a significant relationship in at least one subset, and none yielded conflicting relationships (e.g., a significant positive relationship in one subset and a significant negative relationship in another).

Variables showing positive relationships with effectiveness in all four subsets indicated that the more effective teachers enjoyed teaching and were generally polite and pleasant in their daily interactions. They were more likely to call their students by name, attend carefully to what they said, accept their statements of feeling, praise their successes, and involve them in decision making. This pattern of positive teacher behavior was matched by high ratings of cooperation and work engagement on the part of the students and high ratings on the conviviality of the classroom considered as a whole.

The more effective teachers also were less likely to ignore, belittle, harass, shame, put down, or exclude their students. Their students were less likely to defy or manipulate the teachers. Thus, the classes of more effective teachers were characterized by mutual respect, whereas the classes of less effective teachers sometimes showed evidence of conflict.

The more effective teachers also made demands on students, however. They encouraged them to work hard and take personal responsibility for academic progress, and they monitored that progress carefully and were consistent in following through on directions and demands. Thus, these teachers were pleasant but also businesslike in their interactions with students.

They were also more knowledgeable about their subject matter and effective in structuring it for the students, pacing movement through the curriculum, individualizing instruction, and adjusting to unexpected events or emergent instructional opportunities. They involved all of their students rather than concentrating on a subgroup, and they were more likely to ask open-ended questions and to wait for them to be answered. If aides or other adults were available, these teachers supplemented their own instruction by involving these extra adults in instructional roles.

The more effective teachers were less likely to make management errors such as switching abruptly back and forth between instruction and behavior management, making illogical statements, treating the whole group as one in order to maintain control, and calling attention to themselves for no apparent reason. Finally, they were less likely to kill time with busy work instead of initiating more profitable activities. Taken together, these data indicate that the more effective teachers were more committed to instructing their students in the subject matter, and more knowledgeable, active, and demanding in doing so. They were also better able to match the pace of instruction to the group's needs and to respond to unforeseen events and the needs of individuals. These academic skills were supported by classroom management skills and positive personal characteristics that engendered student attention, task engagement, and general cooperation, resulting in a generally convivial classroom atmosphere.

Several relationships appeared for one grade only (in both subject areas). Teacher and student mobility was greater in the more effective second-grade classrooms. Most likely, this is related to findings reported by others that achievement is lower in classes where students spend a great deal of time working without teacher supervision. The variance in mobility is reduced by fifth grade, when most small-group instruction has been phased out. Several variables were negatively associated with effectiveness only at second grade: expressing distrust of students, publicly verbalizing performance expectations, moralizing, policing, rushing students to answer or finish their work, and overconcern about doing things by the clock. Most of these variables would be expected to correlate negatively with effectiveness measures whenever they did correlate significantly. Use of nonverbal signals to establish control was negatively related to effectiveness in fifth grade. This relationship was not expected, because Kounin (1970) and others have established that nonintrusive control techniques such as nonverbal signaling are usually preferable to more salient techniques that interrupt the flow of instruction. However, the measure recorded the frequency rather than the effectiveness with which such techniques were used, and high frequencies of control attempts suggest deficiencies in more fundamental management skills such as withitness or maintaining signal continuity.

There were two subject matter differences. Teachers' concern about being liked (carried to the extent of trying to ingratiate themselves with students at the expense of instruction) was negatively associated with effectiveness only in mathematics. The reading data were in the same direction, however, and approached significance. Teacher attempts to dispense information and develop positive attitudes about different cultures were positively associated with effectiveness in reading but uncorrelated in mathematics, where there are fewer opportunities to relate the content to cultural differences.

The remaining variables had weaker relationships with effectiveness. Positive relationships were seen for exercising control by praising desirable behavior, defending students from assault, acting as a model, openly admitting mistakes or negative emotions, allowing students to teach one another, and using teacher made materials. Negative relationships were seen for emphasizing competition, using drill activities, differentiating students on the basis of sex, and stereotyping according to SES, race, or ethnicity. None of these findings is surprising except the negative relationship for drill activities, which other investigators sometimes find positively associated with achievement.

The BTES ethnographic data both replicate the major findings from studies using low inference coding and extend those findings in important ways. One major extension is into the affective area. Perhaps better than any others, these data show that academically effective teachers can also be warm, student-oriented individuals who develop a generally positive classroom atmosphere and not merely an efficient learning environment. Concerning instruction, the data indicate the importance of pacing at a rate appropriate to the group and, within this, of responding to the needs of individuals. The following study addressed these instructional issues more specifically.

BTES Phase III-B: Second field study. During 1976-1977, another field study was done in 25 second-grade and 21 fifth-grade classes selected because they contained at least six target students (usually three boys and three girls) whose entry level mathematics and reading scores fell between the 30th and 60th percentiles of the distributions of scores from larger samples of 50 classrooms at each grade level. The result was a racially and ethnically mixed sample weighted toward the lower half of the SES distribution. Except for their willingness to volunteer, the teachers in this study were not preselected, and nothing was known about their relative effectiveness.

Student achievement and attitudes were measured in October, December, and May. The teachers were interviewed at length in the fall and spring, and briefly each week in between. They also kept daily logs. These data were used to assess the teachers' "planning functions" of diagnosis (ability to predict the degree of difficulty that students would experience with particular content) and prescription (allocations of time to various content categories).

Classes were observed for one entire day each week for 20 weeks. Each of the six target students was coded every four minutes for the content being taught, level of attention or task engagement, and apparent level of success (high, moderate, or low). If the teacher happened to be interacting with the student, the teacher's behavior was coded for three "instructional interaction functions" divided into seven categories: presentation (planned explanation of content, unplanned explanation of content, or provision of structuring or directions for tasks), monitoring (observing or questioning the students), and feedback (feedback about academic responses or feedback designed to control attention or task engagement). The data are discussed in technical reports (Berliner, Fisher, Filby, & Marliave, 1978; Fisher et al., 1978) and in a chapter (Fisher et al., 1980) in a larger volume (Denham & Lieberman, 1980) on the BTES Phase III-B findings and their potential policy implications.

Across all classes, only about 58% of the school day was allocated to academics (reading, mathematics, science, social studies), with 24% allocated to nonacademic activities (music, art, story time, sharing), and 18% to non-instructional activities (transitions, waiting, class business). Of the time allocated to academics, students averaged 70-75% actually engaged in academic tasks. They were directly supervised by the teacher only about 30% of the time, spending the other 70% in independent seatwork.

Achievement was associated with the amount of time that students were exposed to academic content (allocated time), the percentage of this time that they actually spent engaged in academic activities (engaged time), and the degree to which they were able to respond to these activities successfully (success rate). Thus, not just the quantity but the quality of student engaged time on task was associated with achievement.

As with the Brophy and Evertson (1974b) data, the findings on success rate varied with context and suggest that different success rates are optimal for different activities and types of student. For the sample as a whole, success rates for individual students averaged almost 50% high success (completely correct work except for occasional, chance level errors due to carelessness), almost 50% medium success (student has general understanding of the task but makes errors at above a chance rate), and only 0-5% low success (student does not understand the task and is able to make correct responses at only a chance rate). Fifth-grade math classes were somewhat more difficult, averaging only about 35% high success rates. Analyses at the individual student level regularly showed negative relationships with achievement for low success rates, and usually showed negative relationships for medium success rates and positive relationships for high success rates. Given the frequencies with which the three success rates were observed, these data imply that high achievement was associated, on the average, with a success rate mixture that approximated 65-75% high success, 25-35% medium success and 0% low success. Either or both of the following causes could explain this association between achievement and a primarily high success rate; high achievers simply make fewer errors than low achievers (student ability effect), or some teachers are better than others at matching instruction and academic tasks to their students' current needs (teacher diagnosis/prescription effect).

Later analyses of these success rate data aggregated to the level of class means (i.e., using the teacher rather than the student as the unit of analysis) suggested that high achievement was associated more with moderate than with high success rates (Burstein, 1980). Here again, however, patterns of relationship varied by context (grade level, subject matter), and interpretation is complicated by the likelihood that teachers whose classes had the highest averages of "high success" time were those who relied most heavily on seatwork and provided less active group instruction to their students.

Taken together, the data suggest that a mixture of high and moderate success rates, with little or no time spent in low success activities, was optimal. High success rates appeared to be more important for younger students (second grade) and for students who had difficulty handling the work. Somewhat more challenge (i.e., moderate success rates) was appropriate for older students (fifth grade).

The BTES authors combined allocated time, engaged time, and success rate into the concept of academic learning time (ALT), which they defined as the time students spent engaged in academic tasks that they could perform with high success. ALT consistently showed significant positive correlations with achievement, and positive but not significant correlations with attitude. Thus these data fit well with other data indicating that high achievement is associated with an instructional pace that is brisk but characterized by gradual movement through small steps with consistent (although not necessarily easy) success, and that a strong academic focus can be achieved without negative effects on student attitudes.

Other positive correlates of achievement included accuracy of diagnosis (ability to predict the difficulty that students would have with particular items), appropriate prescription of tasks (success rates were usually high or moderate, seldom low), frequent provision of academic feedback, emphasis on academic (rather than affective) goals, and student responsibility for academic work and cooperation with academic tasks. Reprimands for misbehavior correlated negatively. Thus classroom organization and management skills and the teaching functions of diagnosis, prescription, and feedback were linked to achievement gain.

Variables connected with the teaching functions of presentation and monitoring did not correlate significantly with achievement, but did correlate with aspects of ALT. In particular, high success rates were associated positively with frequent teacher structuring of lessons and giving of directions for task procedures and negatively with explanations given specifically in response to expressed need. In short, success rates were higher when teachers gave more instruction "up front," before releasing students to work on assignments and less in the form of help for students who had begun assignments but had become confused.

Student engagement rates were associated positively with time spent in "substantive" interaction--when the teacher was giving information about academic content, monitoring work, or giving feedback. Engagement rates were especially low when students spent two-thirds or more of their time working alone.

Teachers who stressed academics elicited the most achievement from students, and teachers who stressed affective objectives elicited the least. The latter teachers not only allocated less time to academics, but showed signs of poor diagnosis and prescription skills. Their classes were more likely to be given tasks that produced low success rates and (therefore?) to show lower task engagement rates. Teachers committed to both academic and affective objectives produced intermediate levels of achievement. Here again one sees that although a strong academic focus can be compatible with positive student attitudes, different objectives ultimately begin to conflict when time allocated in the service of one comes at the expense of time that could be allocated in the service of another.

The BTES Phase III-B data also point up the tension that exists between attempts to maximize student engagement and attempts to maximize success rate. Engagement is generally higher during activities conducted by the teacher than during independent seatwork time. However, group activities expose everyone to the same content and eventually result in moving too slowly for the brightest students but too quickly for the slowest. Differentiated seatwork assignments address this problem by making it possible for all students to achieve at high success rates, but (1) require more teacher preparation and more complex classroom management, (2) result in lower engagement rates despite the increased success rates, and (3) tend to increase the difference between the highest and the lowest achievers in the class. These and other dilemmas raised by BTES Phase III-B data are discussed in the Denhan and Lieberman (1980) volume.

Major contributions of this study are the ALT concept and the demonstration of great variance in allocated time, engaged time, and success rates. Across a school year, some second-grade classes receive an average of 15 minutes of mathematics instruction per day, while others average 50 minutes. Whatever the allocated time, some classes are attentive to lessons or engaged in tasks only about 50% of the time, but others average 90%. Finally, some classes frequently are left to struggle with tasks that are beyond their present abilities, while others rarely are required to endure low success rates, frequently enjoy high success rates, and typically receive sufficient teacher structuring, monitoring, and feedback to enable them to cope effectively with challenging tasks that produce moderate success rates.

Stanford Studies

Throughout the past two decades, Gage and his students and colleagues at Stanford University have been conducting process-product research, especially experimental studies. In the mid 1960s, a series of dissertations (reviewed by Rosenshine, 1968) were designed to study the clarity and effectiveness of teachers' presentations. In each study, teachers were given identical material to teach (suited in difficulty level to their students but not taught as part of the regular curriculum) and asked to present the material during brief (typically 10-minute) time periods. Lessons were videotaped for later analysis, and achievement was assessed with criterion-referenced test scores adjusted for ability.

Fortune (1967) studied student teachers working in Grades 4, 5, or 6 in English, mathematics, or social studies. High inference ratings of teachers' skill in presenting the lesson significantly discriminated between teachers eliciting higher and lower achievement from students in all three subject areas. In addition, five low inference measures of specific teacher behaviors discriminated in two areas, indicating that teachers eliciting higher achievement more frequently (1) introduced the material using an overview or analogy, (2) used review and repetition, (3) praised or repeated pupil answers, (4) were patient in waiting for responses to questions, and (5) integrated such responses into the lesson.

Two other studies used videotapes of experienced 12th-grade social studies teachers' lectures on Thailand and Yugoslavia. One of these, by Rosenshine (described in Gage et al., 1968) involved counting the frequencies of various syntactic, linguistic, and gestural events in the teachers' behavior. Analyses of these codes revealed that the higher achieving teachers used more gestures and movements, more rule-example-rule patterns of discourse, and more explaining links. In the rule-example-rule pattern, the teacher first presents a general rule, then a series of examples, and finally a restatement of the general rule. This contrasts with patterns in which teachers either never state the rule or state it only once rather than giving it both before and after the examples. Explaining links are words that denote cause, means, or purpose: because, in order to, if ... then, therefore, consequently, and so on. By making explicit the relationship between two ideas or events, teachers help insure that students remember the relationship and not merely the ideas or events themselves.

Hiller, Fisher, and Kaess (1969), using transcripts from these same 12th-grade social studies lectures, found that achievement was associated positively with verbal fluency and negatively with vagueness. Vagueness indicators included ambiguous designation (all of this, somewhere), negated intensifiers (not many, not very), approximation (almost, pretty much), "bluffing" and recovery (anyway, of course), error admission (excuse me, not sure), indeterminate qualification (some, a few), multiplicity (sorts, factors), possibility (may, could be), and probability (sometimes, often).

Structuring, soliciting, and reacting. Clark et al., (1979) conducted an experiment in which each of four teachers was trained to teach a nine-lesson ecology unit in eight different ways to eight different randomly assigned groups of sixth graders. The eight different lessons were developed by factorially varying two levels of structuring, two levels of soliciting, and two levels of reacting. High structuring involved reviewing the main ideas and facts covered in the lesson, stating objectives at the beginning, outlining lesson content, signaling transitions between lesson parts, indicating important points, and summarizing parts of lessons as the lessons proceeded. Low structuring involved the absence of these teaching behaviors.

High soliciting was defined as asking approximately 60% higher order questions and 40% lower order questions and waiting at least three seconds for a response after asking a question. Low soliciting involved asking about 15% higher order questions and 85% lower order questions, and calling on a second student to respond if the first did not do so within three seconds. Higher order questions were defined as those requiring mental processes beyond the knowledge level as defined in the *Taxonomy of Educational Objectives* (Bloom et al., 1956).

High reacting involved praising correct responses; negating incorrect responses and giving the reason for the incorrectness; prompting by providing hints when responses were incorrect or incomplete; and writing correct responses on the board. Low reacting consisted of: giving neutral feedback following correct responses; negating incorrect responses but not giving the reason for the incorrectness; and probing or repeating questions following incomplete or incorrect responses, but without giving hints or clues. In all cases, questions were redirected to a second student if probing failed to elicit the correct response from the first; the correct answer was given if neither probing or redirecting elicited it.

Teachers were provided with lesson scripts exemplifying each mixture of instructional components (such as high structuring, low soliciting, and high reacting). Observation indicated that the teachers taught each series of lessons as prescribed and that the lessons did not appear notably different from typical lessons in these classes.

Students were pretested for general abilities and for specific knowledge of the content taught in the unit and were posttested both immediately after the unit and again three weeks later. Testing included attitude measures, an essay test, and a multiple choice test which yielded subscores for higher versus lower order knowledge items and for items that the students could have learned only from the teacher versus from either the teacher or the text. As expected, the treatments showed greater effects on items that had to be learned from the teacher and on lower level knowledge items.

The immediate posttest data showed no effects on the student attitude measure or the essay test. Low soliciting was associated with high scores on both low level and high level items learnable from the teacher only and low level items learnable from either the teacher or the text. In addition to these main effects for low soliciting, there were significant interactions indicating that the combination of low structuring with low reacting yielded low achievement on higher order items learnable only from the teacher and on lower order items learnable from either the text or the teacher. Finally, a nonsignificant trend suggested that high structuring was associated with high achievement on the lower order items learnable only from the teacher.

Data from the retention tests three weeks later were similar. Once again there was no effect on attitude. There was one significant effect for the essay test, however, indicating that high scores were associated with high reacting. In addition, scores for lower order multiple choice items learnable only from the teacher were associated with high structuring, low soliciting, and high reacting. Also, interaction effects again indicated that the combination of low structuring and low reacting was particularly dysfunctional.

In general, these data support other findings indicating the importance of teachers' structuring the content through clear presentations, providing feedback to student responses, and attempting to improve responses that are incomplete or incorrect, and indicating that a predominance of lower order questions is associated with high achievement gain, even on items dealing with higher order content.

Program on teaching effectiveness. More recently, Gage and his colleagues in the Program on Teaching Effectiveness at Stanford University have conducted two additional studies involving training teachers to implement 22 principles suggested by 81 findings reported by others. Approximately 50% of these findings were drawn from Brophy and Evertson (1974a,1974b), 31% from Stallings and Kaskowitz (1974) 15% from McDonald and Elias (1976b) and 4% from Soar (1973). Some principles were intended for use with all students, but others were targeted for students described as either "more academically oriented" (high achieving, well motivated) or "less academically oriented" (low achieving, possibly anxious or uncooperative).

Third-grade teachers working in middle SES schools were first stratified according to mean academic achievement of their students, then randomly assigned to three groups: observation only (N = 10), minimal training plus observation (N = 11), or maximal training plus observation (N = 12). Minimally trained teachers were merely mailed packets discussing the principles (one packet per week for five weeks). Maximally trained teachers received the packets at the same rate, but also participated in a two hour meeting each week to discuss the recommendations. Classes in all three groups were observed for four full days prior to the treatment, another four or five days during November and December after the teachers received the packets, and another seven days between January and May. Analyses indicated that about half of the training components were implemented successfully and that the means for the experimental groups typically were nearer to the prescribed guidelines than the means for the control group. Unexpectedly, the minimal training group implemented the guidelines somewhat better than the maximal training group.

Adjusted achievement in vocabulary for the combined treatment groups exceeded that of the control group by 0.69 standard deviation units, which approached but did not reach statistical significance ($p<0.15$). There was no comparable effect on reading comprehension. Process-product correlations based on the total sample of 33 teachers supported the findings reported in earlier studies only about half of the time. Much of this agreement was with Brophy and Evertson's findings for high SES students comparable to those included in the present study (Crawford et al., 1978). Once again one sees the need to consider student SES in interpreting process-product data from the early grades, particularly data on reading instruction.

A variation of this experiment was repeated in a subsequent study of 28 classes in fourth through sixth grades in a school serving a low SES, predominantly black population (Gage & Coladarci, 1980). All teachers experienced "minimal" training (receiving one packet per week for five weeks, by mail) but without any personal contact with the experimenters. Classrooms were also observed, but only for 2 two-hour observations before and again after the treatment. This time, implementation was poor: the training related behaviors of the 15 experimental teachers were not altered appreciably by the treatment and did not differ significantly from those of the 13 control teachers. Nor did the achievement of treatment classes exceed that of control classes.

Despite this lack of treatment effect, process-product data based on the total sample of 28 classes indicated that the teacher behaviors, particularly those related to classroom management and time spent in academic activities, called for in the guidelines were correlated with achievement as expected. These relationships were strongest in the fourth grade and weakest in the sixth grade, which was to be expected because the guidelines were based on data from the primary grades. Phonics instruction, which typically correlates positively with reading achievement in the early grades, correlated negatively in these middle grades.

Clarity Studies

The work of Rosenshine (1968) and of Hiller, Fisher, and Kaess (1969) on clarity of teacher presentations has been elaborated in recent years. Issues of definition and measurement have been discussed by McCaleb and White (1980), Cruickshank, Kennedy, Bush, and Myers (1979) and Kennedy, Cruickshank, Bush, and Myers (1978). In addition, Land and Smith and their colleagues have contributed a dozen individual studies and reviews (summarized in Smith & Land, 1981) concerning relationships between low inference measures of teacher clarity and achievement. Most of these have been conducted with college students as subjects, although junior high and high school studies have been included. Typically, groups of students are randomly assigned to listen to and then take a test on an audiotaped lesson. Different versions of the lesson are prepared by varying the presence of elements that detract from clarity. The most commonly studied of these are the "vagueness terms" described by Hiller et al. (1969). Smith and Land (1981) report that adding vagueness terms to otherwise identical presentations reduced student achievement in all 10 of the studies in which vagueness was manipulated. Vagueness terms are italicized in the following excerpt:

> This mathematics lesson *might* enable you to understand a *little more* about *some things* we *usually* call number patterns. *Maybe* before we get to *probably* the main idea of the lesson, you should review *a few* prerequisite concepts. *Actually*, the first concept you need to review is positive integers. *As you know*, a positive integer is any whole number greater than zero. (Smith & Land, 1981, p. 38)

Clarity can also be reduced by "mazes," which are false starts or halts in speech, redundantly spoken words, or tangles of words. Inclusion of mazes in presentations reduced achievement in three of four studies. Mazes are italicized in the following excerpt:

> This mathematics lesson *will enab* ... will get you to understand *number, uh*, number patterns. Before we get to the *main idea of the*, main idea of the lesson, you need to review *four conc* ... four prerequisite concepts. The first *idea, I mean, um*, concept you need to review is positive integers. A positive *number* ... integer is any whole *integer, uh*, number greater than zero. (Smith & Land, 1981, p. 38)

A third element that can detract from clarity is discontinuity, in which the teacher interrupts the flow of the lesson by interjecting irrelevant content or by mentioning relevant content at inappropriate times. Kounin

(1970) included such discontinuities among reasons for loss of lesson momentum. More recently, Land and Smith (1979) found that extra content interjected into presentations did not affect achievement, but Smith and Cotten (1980) found that interjected discontinuities significantly reduced achievement. The latter study involved more drastic changes from the original clear presentation, which probably accounts for the difference in results.

A fourth detractor from clarity is saying "uh." This had a negative but nonsignificant relationship with achievement in the one study in which it was investigated in its own right (Smith, 1977). It also has been included along with the other three detractors (vagueness terms, mazes, and extra content) in studies that used a cluster of six variables to create high and low clarity treatments. Two positive elements in these clusters were emphasis on key aspects of the content to be learned and clear signaling of transitions between parts of lessons. Lessons constructed to maximize clarity by including these positive elements and avoiding the detractors discussed above typically produce greater achievement than less clear lessons (Land, 1979).

Other aspects of clarity, such as structuring and sequencing the content and explaining it understandably (see McCaleb & White, 1980) have been addressed by other researchers (even though not all of them use the term "clarity" in describing their data). In general, clarity of presentation is one of the more consistent correlates of achievement, at least in studies where exposure to the content to be tested is controlled.

Additional Studies

We have reviewed the programmatic work of several teams of investigators. Before initiating integrative discussion, we conclude the review with brief summaries of additional studies that meet the inclusion criteria stated at the beginning of the chapter.

Correlational Studies

Several correlational studies linked achievement to opportunity to learn the content included on tests. Content coverage was measured directly by asking teachers to state whether or not (or how much) they covered specified content (Borg, 1979; Chang & Raths, 1971; Comber & Keeves, 1973; Harris & Serwer, 1966; Husen, 1967), by coding the content relevance of classroom activities and questions (Smith, 1979), or by doing both (Cooley & Leinhardt, 1980). Other studies documented the same relationship indirectly by relating achievement to the percentages of time spent in academic activities rather than procedural or disciplinary interactions (Dalton & Willcocks, 1983; Emmer, Evertson, & Anderson, 1980; Evertson & Emmer, 1982; Fitz-Gibbon & Clark, 1982; Galton & Simon, 1980; Rose & Medway, 1981).

These "opportunity to learn" findings are sometimes also described as "time allocation" or "time on task" findings. The latter terms are less desirable because they are less accurate and specific (Borg, 1980). Furthermore, they require at least three qualifications. First, the data indicate the need to consider the quality of academic activities and not just the time spent on them. Fisher et al. (1980) elaborate this point in discussing the BTES Phase III-B data. Second, the time on task that is linked most closely to achievement is time spent in teacher directed lessons or in seatwork actively supervised by the teacher. Large amounts of time spent working without supervision are associated with low achievement gain. Finally, although measures of time allocation to academic activity (and especially measures of time spent actually engaged in those activities) typically correlate positively with achievement (Borg, 1980), these correlations are usually only weak to moderate (Fitz-Gibbon & Clark, 1982; Karweit, 1983), and they vary according to the definition and measurement of time on task (Karweit & Slavin, 1982).

Thus, efforts to determine the implications of research on teacher effects should concentrate on issues of opportunity to learn and quality of instruction. Time on task does not translate into achievement in any simple or direct way (Brophy, 1979; Karweit, 1983; Wyne & Stuck, 1982).

Arehart. Arehart (1979) studied 23 teachers who taught a three-period probability unit to 26 classes in 8th through 11th grade. All teachers taught to the same objectives using the same content outline and problem exercises but were free to teach in their own way. Achievement correlates included content covered by the teacher; percentage of assigned problems attempted by the students; and percentages of total interaction classified as "substantive," as teacher informing, and as teacher questioning. There was no significant relationship with achievement for pupil initiations. In general, percentage measures correlated more strongly than frequency measures, and teacher informing measures more strongly than teacher questioning measures. Teacher informing did not necessarily mean extended lecturing, however. More typically, it involved giving information for a minute or less and then asking a question.

Armento. Armento (1977) studied 20 preservice and 2 inservice teachers who delivered social studies lessons to students in third through fifth grades. High inference correlates of achievement included ratings of accuracy of examples, relevance of teacher behavior to learning objectives, balance between concrete and abstract terminology, and expression of interest and enthusiasm. Low inference correlates included giving definitions, examples, and labels for concepts; summarizing or reviewing main ideas; and general adequacy of content coverage. No significant relationships appeared for signaling changes in topic, asking questions (either lower order or higher order), repeating or rephrasing questions, asking questions in pairs, or telling students to stop irrelevant behavior.

Boak and Conklin. Boak and Conklin (1975) studied 10 mathematics teachers in seventh through ninth grade and 20 language arts teachers in seventh and eighth grade. The teachers were classified as either high or low in interpersonal skills (based on ratings of empathy, respect, and genuineness developed from audiotaped lesson segments and on ratings of empathy based on their written responses to vignettes depicting student concerns). These classifications were then related to achievement gain. There was no relationship in the seventh grade language arts classes, but the students of the higher rated teachers made greater gains in reading comprehension in the eighth-grade language arts classes, and in mathematics in the mathematics classes. Although the teacher classifications in this study were not based solely on observation of classroom behavior, the data suggest that the interpersonal skills stressed by Aspy (1969,1972), Carkhuff (1969,1971) and others may correlate with achievement in addition to affective outcomes.

Coker, Medley, and Soar. Coker, Medley, and Soar (1980) reported process-product data from 100 classes in 1st through 12th grade, 59 studied the first year and 41 the second year. The findings are difficult to evaluate because they are reported only for the sample as a whole rather than separately by grade levels and because the process measures are combination scores that include data from both academic and non-academic activities. Still, a few general trends are discernible in the correlations that were significant in both years. Positive correlates of achievement included selecting appropriate goals and objectives for students; involving the students in organizing and planning; giving clear, explicit directions; and listening to students and respecting their right to speak during recitations and discussions. Negative correlates included poor classroom management, overemphasis on praise and rewards (probably also related to poor classroom management), overemphasis on eliciting and responding to student questions (perhaps reflecting insufficient or ineffective presentation of information by the teacher), and overemphasis on student input into decision making. The latter findings seem reminiscent of the BTES Phase III findings suggesting that teachers who concentrated either on affective objectives or on ingratiating themselves with their students produced less achievement than teachers who concentrated on cognitive objectives.

Crawford. Crawford (1983) studied instruction in 79 first- through eighth-grade compensatory education classes for Title I students. These classes were small (5-10 students), intended to remediate weaknesses in basic reading and mathematics skills, and taught by specially trained teachers assisted by paraprofessional aides. Across grade level and subject matter, achievement gain was associated with allocation of high percentages of available time to academic activities, good monitoring and other classroom management techniques that maximized task engagement and minimized interruptions and transition time, and active instruction of the students. Much of this instruction was accomplished through interactions with individuals in teaching reading in the early grades, but instruction usually occurred with groups in upper grade reading classes and in mathematics classes at all grade levels. Success in the early grades (in both subject areas) was also associated with teachers' academic demands on students in the form of challenging assignments and frequent attempts to improve initially unsatisfactory answers to questions. The findings that primary grade reading achievement was associated with challenging assignments and with individualized instruction rather than group lessons contrast with most other findings for low SES or low ability students in these grades. They indicate that methods that are impractical in ordinary classes can be used effectively in special classes with small student-teacher ratios. In particular, teachers can move students through curricula at a faster pace, can provide more tutorial and individualized instruction, and can assign more difficult seatwork when the number of students in the class is small enough to allow them to consistently monitor everyone's progress and provide help when needed.

Dunkin. Dunkin (1978) studied 29 sixth-grade teachers asked to teach 30-minute discussion lessons in social studies. Achievement correlates included content coverage, structuring (number of teacher structuring moves as defined by Bellack, Hyman, Smith, & Kliebard, 1966), percentage of total academic questions that were higher order (the average was only 25%), number of relevant pupil responses to teacher questions, and percentage of teacher reactions to student responses that were positive (praise) reactions (these averaged 16%). In addition, there was a nonsignificant negative trend for frequency of teacher vagueness terms.

Dunkin and Doenau. Dunkin and Doenau (1980) studied most of the same teachers studied earlier by Dunkin (1978), this time teaching two additional social studies lessons (*N* = 28 for lesson one, 26 for lesson two). Achievement correlates in both lessons included content coverage through teacher informing statements, content coverage through teacher-student interaction, and total content coverage. Several variables correlated significantly in one lesson but not the other. Of these, positive correlates included total content repetition, percentage of total student words that were classified as vague, asking multiple secondary questions, student initiations that were not questions, student responses that were rejected, and long student utterances. Negative correlates for one lesson only were percentage of informing elements that were terminal rather than initial or intervening, percentage of total questions that were higher order (in this case, the average was 36%), frequency of student initiated questions, frequency of positive reactions to student responses, and total teacher reactions to student responses. There was a nonsignificant negative trend for teacher vagueness in each lesson. Of the variables considered by Dunkin (1978) and by Dunkin and Doenau (1980), consistent relationships with achievement were found only for content covered and (less strongly) for teacher vagueness. Variables connected with teacher structuring, soliciting, and reacting or with types of pupil participation did not yield consistent patterns.

Larrivee and Algina. Larrivee and Algina (1983) observed in 118 elementary grade (K-6) classes that each contained a special education student who was being mainstreamed and would be present during reading and language arts instruction (most of these mainstreamed students were classified as learning disabled). Classrooms were observed four times with each of four instruments, concentrating on the mainstreamed student and certain other target students. The mainstreamed students' reading achievement was associated positively with higher ratings of teachers for efficient use of time, good relationships with students, supportive response to low ability students, and high frequency of positive feedback to student performance. Negative correlates included the frequency of interventions concerning misconduct, time spent off task, and time spent in transitions. Variables that correlated with academic learning time, although not significantly with reading achievement, included the frequency of easy questions, correct student responses, and attempts to improve incorrect responses. In general, these data on achievement correlates for mainstreamed special students parallel the findings for low SES and low achieving students in other elementary grade studies.

McConnell. McConnell (1977) related high inference measures of teacher behavior to student attitude and achievement in 43 ninth-grade algebra classes. Positive attitudes were associated with teacher clarity, enthusiasm, and task orientation. Negative attitudes were most likely in classes that emphasized analysis (such classes were seen as harder and duller; perhaps the teachers were generally low on clarity and enthusiasm). Achievement in both computation and comprehension was correlated with teacher task orientation. Achievement in comprehension was also correlated with clarity, and achievement in analysis (the most abstract measure) was correlated with probing, enthusiasm, and teacher talk.

Solomon and Kendall. Solomon and Kendall (1979) studied 50 fourth-grade classes in relatively affluent schools. They focused on interactions between teacher types and student types and reported most data in terms of combination scores. However, they noted a main effect indicating that classes rated as controlled and orderly showed greater achievement than less controlled or disorganized classes. Interaction effects indicated that low SES students did best in warm, encouraging classrooms, but high SES students did best in more impersonal and academically demanding classrooms. Also, students who preferred autonomy generally did better in the more controlled classes, whereas those who preferred structure generally did better in the more permissive classes (differences between what is preferred and what maximizes achievement have also been reported by others; see Clark, 1982). Other data revealed additional interaction effects and contrasts between what correlated with achievement and what correlated with other outcomes (attitudes, motivation, creativity).

Experimental Studies

Alexander, Frankiewicz, and Williams. Alexander, Frankiewicz, and Williams (1979) studied five variations of social studies lessons taught to fifth- through seventh-grade students. Control students were taught for 50 minutes without the use of organizers as described by Ausubel (1968). In the four experimental treatments, 10 of the 50 minutes were allocated to presentation of superordinate concepts under which the more specific material could be subsumed. In various treatments, the organizers were either visual (photographic slides) or oral-interactive (presentation followed by structured discussion) and were placed either before or after the rest of the lesson. All four organizer groups retained more content than the control group, but no organizer group differed significantly from any of the others. Most tests of Ausubel's ideas about organizers have involved advance organizers included in written materials prepared for independent study by high school or college students. The present study has shown that organizers designed to help students structure their learning can (1) facilitate achievement even at the elementary level, (2) take visual or oral-interactive form in addition to written form, and (3) be effective when placed after the body of the lesson as post organizers (not just prior to it as advance organizers).

Bettencourt, Gillett, Gall and Hull. Bettencourt et al. (1983) studied the effects of enthusiasm training in two studies involving beginning teachers in the elementary grades. The training was effective in each study in that the trained teachers were rated as more enthusiastic than control teachers in their instruction during a special experimental unit on probability and graphing. Effects on outcomes, however, were mixed. Student on-task behavior was the outcome measured in one study, and the treatment produced higher on-task percentages, not only in teacher directed activities but also during seatwork times. However, achievement was the outcome measure in the second study, and this time the data revealed no significant differences. Thus, the enthusiasm training produced some desirable effects, but these were not strong enough to increase achievement significantly.

Blaney. Blaney (1983) essentially replicated the aspects of the Clark et al. (1979) study that dealt with teacher structuring and reacting. A single trained teacher taught four versions (high structuring/high reacting, high structuring/low reacting, low structuring/high reacting, low structuring/low reacting) of the same four-day sequence of science lessons to groups of second graders using semiscripted lessons to control content coverage. Lessons involving high structuring were longer than those involving low structuring, but level of structuring nevertheless was unrelated to achievement. Reacting was related, however; high reacting produced higher achievement.

Clasen. Clasen (1983) studied the effect of four different presentations (independent study, 75% low level questions, 75% high level questions, or 75% divergent production questions) of identical content on the achievement of gifted seventh graders in week-long science units. These contrasting treatments did not make much difference, possibly because all of the students were gifted and thus likely to learn the material if given the opportunity to do so. The lower order question group outperformed both the independent study group and the higher order question group on the lower order items included in an immediate posttest. These group differences disappeared, however, on a delayed retention test. There were no group differences in either test for higher order content items or divergent production items. The student attitude inventory revealed that students in the divergent production group had more positive attitudes toward their experiences during the unit than did students in the independent study group. There were no other group differences.

Gall, Ward, Berliner, Cahen, Winne, Elashoff, and Stanton. Gall et al., (1978) studied the effects of varying recitation and questioning techniques on sixth-grade students' achievement following specially prepared two-week ecology units. In the first study, three groups were taught using 15 minutes of content presentation followed by 25 minutes of recitation. A fourth group (no recitation) engaged in ecology-related art activities following the content presentation. Within the three recitation groups, there was variation in probing (asking follow up questions to try to improve an initial answer) and redirection (calling on another student to respond to a question answered by the first student). The three recitation groups learned more than the art activity group, but there was no evidence that recitations involving probing and redirection were superior to recitations that did not include these elements.

In the second study, the recitation treatments differed in cognitive level of questions asked. One group received 25% higher level questions, the second group 50% and the third group 75%. Once again, the three recitation groups outperformed the art activity group. The results for level of question were puzzling because the 50% higher cognitive level question treatment was less effective than the other two for promoting acquisition and retention of facts, but slightly more effective for promoting performance on higher cognitive level tasks. The scores for the 75% group were similar to, but lower than, those for the 25% group, even on higher cognitive level measures. Taken together, these two studies suggest that students benefit from recitations that allow them to answer questions about content previously presented by teachers, but do not support the hypothesized benefits of probing, redirection, or higher level questions.

MacKay. MacKay (1979) studied third- and sixth-grade classes in Edmonton, Canada. Teachers were trained on 28 strategies drawn mostly from previous process-product research (a few strategies were included because they had been recommended by curriculum specialists). All teachers were observed prior to the treatments, then exposed to the treatments (either two or four half days of inservice activities) and observed again. The treatments produced significant increases for 24 of the 28 strategies, suggesting generally good implementation. Process-product data showed no significant relationships in third-grade reading, where there was very low variance in adjusted achievement scores. However, 16 of the 28 strategies showed significant process-product relationships in the third-grade mathematics classes, 9 in the sixth-grade reading classes, and 2 in the sixth-grade mathematics classes. This pattern of significant relationships was spotty, but all significant relationships were in the expected direction. Most of these involved classroom organization, group management, and responsiveness to students' answers to questions. (The correlates unique to second-grade math included teacher acceptance and caring, academic learning time, interest value of assignments, and checking of seatwork performance.)

McKenzie and Henry. McKenzie and Henry (1979) developed experimental support for an innovation designed to make teachers' yes-no questions function as "test-like events" rather than mere "nominal stimuli" to each student in the class (not just to the student called on to respond). Third graders were randomly assigned to lessons with standardized content presentation and follow up questions. In the control classes, individual students were called on (randomly) to answer questions while their classmates looked on (this is most teachers' typical recitation procedure). In the experimental class, however, all students were required to respond to every question (using nonverbal gestures). This approach reduced off-task behavior and increased achievement.

Madike. Madike (1980) assigned student teachers to teach five-week mathematics units to comparable ninth-grade classes. One group of student teachers had been trained in teaching skills through a microteaching program. A second group had been observed and given feedback by supervising teachers, but not necessarily on the skills stressed in the microteaching program. A third group was given no specific preparation for the teaching experience. Each student teacher was videotaped during a 35-minute lesson, and a 10-minute segment was rated for frequency of use of nine skills taught in the microteaching program. The microteaching group had higher frequencies of behaviors related to these nine skills, and the skills correlated positively (as expected) with achievement. Correlations were significant for questioning, closure (structuring at the ends of episodes initiated by questions), and cuing (verbally calling attention to important content), but not for stimulation, variation, reinforcement, planned repetition, recognizing student attention, using examples, or nonverbal cuing. Even though these data are frequency scores for Nigerian student teachers and were developed from a very limited observation base, they correspond well with other data reviewed here.

Martin. Martin (1979) used an intensive reversal design and time series analyses to assess the impact of increases in higher order questions during an experimental biology unit taught in a sixth-grade class. A baseline period in which the teacher taught normally was followed by an initial experimental phase in which the teacher increased the frequency of higher order questions, then by a return to baseline, and then by another increase in higher order questions. During each phase, teacher questioning and student responding were monitored, and student achievement and attitudes were measured. Results indicated that increases in higher order questions led to increases in higher order responses. However, there was no effect on achievement or attitudes toward lessons and a negative effect on attitudes toward the teacher. Thus the treatment produced the intended changes in processes, but not in outcomes.

Ryan. Ryan (1973,1974) conducted two studies of the effects of level of question in lessons taught to fifth and six graders during inquiry-oriented social science lessons. Each study involved two recitation/discussion groups and a control group that received lectures and completed assignments but were not involved in recitation/discussion activities or in the special activities included in the inquiry program. One discussion/recitation group received about 75% high level questions, and the second received only about 5% high level questions. The inquiry/recitation/discussion groups usually outperformed the control group on both low and high level objectives, but they never differed significantly from each other. Thus, both high and low level questions were effective in promoting achievement of both high and low level objectives.

Schuck. Schuck (1981) studied the effects of set induction on learning in ninth-grade biology lessons. All teachers used the same materials to teach to the same objectives, but the experimental teachers began by inducing a learning set by drawing analogies between the new material and events that were already familiar to the students. Students exposed to the set induction treatment learned and retained more content than control students.

Smith and Sanders. Smith and Sanders (1981) studied the effect of high versus low structuring of fifth-grade social studies content. Following Anderson (1969), they defined structure in terms of linear redundancy in the appearance of key concepts. In high structure presentations, key concepts tend to be repeated from one sentence to the next, although new ones are gradually phased in and old ones phased out. This structure is typical of prose that moves systematically through a series of related statements. In low structure presentations, the content was more jumbled. Key concepts were repeated just as often, but not in contiguous sentences. As a result, even though the same sentences were included in each version, the high structure presentations were clearly recognizable as organized sequences of related facts, but the low structure presentations sounded more like lists of unrelated facts. As expected, the high structure presentations produced higher student achievement and ratings of effectiveness.

Tobin. Tobin (1980) studied the effect of increasing "wait-time" on learning in science classes for Australian students aged 10 to 13. Tobin's definition of wait-time was considerably broader than the definition used by Rowe (1974) in her investigations of the effects of pausing for several seconds after asking a question (in order to give the students time to think about the question before calling on one of them to try to answer it).

Tobin's definition of wait-time included not only these pauses, but also teacher pauses following student responses or previous statements by the teacher. Thus, wait-time was defined by Tobin as the length of the pause preceding any teacher utterance.

Prior to treatment, the mean wait-time for all teachers was 0.5 seconds. Following treatment, these averages were 3.1 for the experimental teachers and 0.7 for the control teachers. There was no significant correlation between wait-time and achievement before the treatment (probably because there was no meaningful variation in wait-time), but a positive correlation afterwards. This was true even though only 8 of the 13 experimental teachers succeeded in meeting the criterion of an average wait-time of three seconds, and some of them did not view such long wait-time as appropriate. It should be noted that these lessons involved scientific concepts such as density and displacement. Such extended wait-time might be less appropriate in lessons involving simpler content or younger students.

Tobin and Capie. Tobin and Capie (1982) manipulated both teacher wait-time and quality of questioning (cognitive level, clarity, relevance) in middle school science lessons. There were four groups: (1) extended teacher wait-time plus high question quality, (2) entended teacher wait-time plus normal question quality, (3) normal teacher wait-time plus high question quality, (4) normal teacher wait-time plus normal question quality. Teachers in the extended wait-time groups were asked to average between three and five seconds of wait-time (again, wait-time was defined as the length of time preceding a teacher utterance). Teachers in the high question quality groups were asked to plan their questioning to be high in cognitive level, clarity, and relevance (which included both relevance to the objectives of the lesson and appropriateness of timing given the flow of the lesson). The teachers were observed and given feedback to help them maintain the specified wait-time and question quality levels.

Wait-time showed a significant positive correlation with achievement, and there were positive but nonsignificant relationships for cognitive level, clarity, and relevance of questioning (variance in question quality was low, because all teachers tended to ask questions that were high in cognitive level, clarity, and relevance, because they all were given detailed lesson plans).

Although wait-time correlated positively with achievement, it also interacted with question quality and showed a curvilinear relationship to student engagement. The interactions suggest that longer wait-times are especially important when instruction deals with higher cognitive level objectives, and that a mix of questions at varying cognitive levels produces the highest achievement (a ratio of approximately two higher level questions to one lower level question was optimal in these data). The highest rates of attending were associated with wait-times of approximately three seconds (as opposed to shorter or longer wait-times) combined with intermediate cognitive levels of question (as opposed to lower or higher levels).

Summary and Integration of the Findings

Earlier *Handbooks'* chapters on teacher effects concentrated on issues of definition and methodology, because there were few replicated findings to discuss. However, research of the 1960s and 1970s yielded numerous replicated linkages between teacher behavior and achievement. Many of these linkages have even been validated experimentally, although it remains true that experimental findings are weaker and less consistent than correlational findings.

The emphasis here is on consistency and replication of findings, not size of correlation. Even the most generally replicated findings tend to be based on low to moderate correlations, and many findings must be qualified by reference to grade level, student characteristics, or teaching objectives. This reflects the fact that effective instruction involves selecting (from a larger repertoire) and orchestrating those teaching behaviors that are appropriate to the context and to the teachers' goals, rather than mastering and consistently applying a few generic teaching skills.

Research based conclusions about teacher behaviors that maximize student achievement are summarized below, first for general aspects of instruction and then for the handling of specific lesson components. The evidence supporting these conclusions is strongest for basic skills instruction in the primary grades, but extant data suggest that they also apply to instruction in certain subjects at all grade levels (limits and qualifications on the data are discussed in the next major section).

Quantity and Pacing of Instruction

The most consistently replicated findings link achievement to the quantity and pacing of instruction.

Opportunity to learn/content covered. Amount learned is related to opportunity to learn, whether measured in terms of pages of curriculum covered or percentage of test items taught through lecture or recitation. Opportunity to learn is determined in part by length of school day and school year, and in part by the variables discussed below.

Role definition/expectations/time allocation. Achievement is maximized when teachers emphasize academic instruction as a major part of their own role, expect their students to master the curriculum, and allocate most of the available time to curriculum related activities. This is seen in relationships involving presage measures of teachers' role definition and expectations, high inference ratings of the degree to which teachers are businesslike or task oriented, and low inference measures of time allocated to academic activities rather than to activities with other objectives (personal adjustment, group dynamics) or with no clear objectives at all ("free time," student choice of games or pastimes).

Classroom management/student engaged time. Not all time allocated to academic activities is actually spent engaged in these activities. Engagement rates depend on the teacher's ability to organize and manage the classroom as an efficient learning environment where academic activities run smoothly, transitions are brief and orderly, and little time is spent getting organized or dealing with inattention or resistance. Key indicators of effective management include (1) good preparation of the classroom and installation of rules and procedures at the beginning of the year, (2) withitness and overlapping in general interaction with students, (3) smoothness and momentum in lesson pacing, (4) variety and appropriate level of challenge in assignments, (5) consistent accountability procedures and follow up concerning seatwork, and (6) clarity about when and how students can get help and about what options are available when they finish (see Chapter 16 of the *Handbook*).

Consistent success/academic learning time. To learn efficiently, students must be engaged in activities that are appropriate in level of difficulty and otherwise suited to their current achievement levels and needs. It is important not only to maximize content coverage by pacing the students briskly through the curriculum, but also to see that they make continuous progress all along the way, moving through small steps with high (or at least moderate) rates of success and minimal confusion or frustration. If lessons are to run smoothly without loss of momentum and students are to work on assignments with high levels of success, teachers must be effective in diagnosing learning needs and prescribing appropriate activities. Their questions must usually (about 75% of the time) yield correct answers and seldom yield no response at all; their seatwork activities must be completed with 90-100% success by most students.

(Such high success rates should not be taken as suggestive of instructional overkill or assignment of pointless busy work. Appropriate seatwork will extend knowledge and provide needed practice. It will also be "doable," however, because it is pitched at the right level and the students have been prepared for it. Thus the high success rates result from effort and thought, not mere automatic application of already overlearned algorithms.)

Continuous progress at high rates of success, carried to the point that performance objectives can be met smoothly and rapidly, is especially important in the early grades and whenever students are learning basic knowledge or skills that will be applied later in higher level activities.

In summary, then, there is a tension between the goal of maximizing content coverage by pacing the students through the curriculum as rapidly as possible and the needs to (1) move in small steps so that each new objective can be learned readily and without frustration, (2) see that the students practice the new learning until they achieve consolidated mastery marked by consistently smooth and correct responses, and (3) where necessary, see that the students learn to integrate the new learning with other concepts and skills and to apply it efficiently in problem solving situations. The pace at which the class can move will depend on the students' abilities and developmental levels, the nature of the subject matter, the student/teacher ratio, and the teacher's managerial and instructional skills. In general, teachers should hold errors to a minimum by choosing tasks that their students can handle and explaining those tasks clearly before releasing the students to work on them. The more challenging the task, the more the teacher must be prepared to monitor performance as the students work on the task (not just to correct answers later) and to provide immediate help to those who need it.

Bennett, Desforges, Cockburn, and Wilkenson (1981) point out that not only the frequency of errors is important, but their timing and quality. Early in a unit, where new learning is occurring, relatively frequent errors may be expected. Later, however, when mastery levels are supposed to have been achieved, errors should be minimal. Also, some errors occur because students have the right general idea but make a minor miscalculation, or because they involve sound logic that is based on assumptions that are plausible but happen to be faulty. Such "high quality" errors are understandable and may even provide helpful guidance to the teacher. However, errors that suggest inattention, hopeless confusion, or alienation from the material are undesirable.

Active teaching. Students achieve more in classes where they spend most of their time being taught or supervised by their teachers rather than working on their own (or not working at all). These classes include frequent lessons (whole class or small group, depending on grade level and subject matter) in which the teacher presents information and develops concepts through lecture and demonstration, elaborates this information in the feedback given following responses to recitation or discussion questions, prepares the students for follow up seatwork activities by giving instructions and going through practice examples, monitors progress on assignments after releasing the students to work independently, and follows up with appropriate feedback and reteaching when necessary. The teacher carries the content to the students personally rather than depending on the curriculum materials to do so, but conveys information mostly in brief presentations followed by recitation or application opportunities. There is a great deal of teacher talk, but most of it is academic rather than procedural or managerial, and much of it involves asking questions and giving feedback rather than extended lecturing.

The findings just summarized all deal with quantity of academic activity, particularly the time spent in organized lessons and supervised seatwork. The following variables concern the form and quality of teachers' organized lessons.

Whole Class versus Small Group versus Individualized Instruction

The data do not say much about teaching the whole class versus small groups. No experimental studies have compared these two lesson formats directly, and the issue was not addressed correlationally except in the Follow Through studies where it was confounded with other systematic differences. Even in the absence of definitive data, certain trade-offs are obvious. Whole class instruction is simpler in that the teacher needs to plan only one set of lessons and is free to circulate during seatwork times (although teaching the whole class is more demanding than teaching any particular small group). The small-group approach involves preparing differentiated lessons and assignments and keeps the teacher busy instructing small groups most of the time (and thus unavailable to monitor and assist the majority of students who are working on assignments). The small-group approach, then, requires well chosen assignments that the students are willing to engage in and able to complete successfully, as well as rules and procedures that enable students to get help (if confused) or direction (about what to do when finished) without disrupting the momentum of the teacher's small-group lessons. Unless they have an aide, even teachers who are able to make the small-group approach work may find that it takes too much effort to be worth the trouble.

However, small-group instruction may be necessary in at least two situations. The first is beginning reading instruction, where it is essential that each individual read aloud so that the teacher can monitor progress and diagnose and correct consistent error patterns. The slow pace, repetition, and sustained attention to individuals that such instruction requires are incompatible with the brisk pacing that makes for successful whole-class lessons. Grouping (although not necessarily ability grouping) is a way for teachers to accommodate the slow paced reading turns that characterize beginning reading instruction. It can be phased out as reading lessons evolve from decoding to comprehension objectives.

Grouping may also be necessary in highly heterogeneous classes. Here, grouping may be based on differences in ability, achievement, or language dominance, and different groups may receive both different instruction and different assignments. This requires more complex planning and group management than whole-class instruction and introduces the potential for undesirable expectation or labeling effects, but there may be no alternative in many classes.

It should be noted that these remarks about grouping refer to the tradeoffs involved in differentiating the class to allow for separate instruction or assignments. They do not apply to the use of student teams, tournaments, and other approaches that Slavin (1980, 1983) and others have recommended for boosting motivation and increasing pro-social peer contact. These approaches involve introducing cooperative or competitive (not merely individualistic) reward structures to the management of seatwork, and can be used with whole-class, small-group, or individualized instruction.

The studies reviewed here do not have much to say about individualized instruction, because process-outcome researchers have concentrated on teacher-led instruction (other reviews suggest mixed findings; see Good & Brophy, 1984). In particular, these data are silent on the relative merits of specific programs of individualized instruction (Individually Guided Education, Individually Prescribed Instruction, etc.). However, they do show consistent positive correlations with achievement for active (whole-class or small-group) instruction by the teacher, and negative correlations for time spent in independent seatwork without continuing teacher supervision. Thus, although these data do not contradict the notion of individualizing instruction as a general principle, they do raise doubts about the probable effectiveness of particular programs of individualized instruction in which students are expected to learn mostly on their own from reading curriculum materials, working on assignments, and taking tests. This approach to individualized instruction does not appear feasible in ordinary classes, although it can work in special classes with low student-teacher ratios (c.f. Crawford, 1983).

In summary, small-group instruction is more complex to implement than whole-class instruction, but it may sometimes be necessary. Available data are not very informative about when small-group instruction should be considered the method of choice, nor about how it should be designed and managed. "Individualized instruction" which relies heavily on unsupervised independent seatwork is not as effective as teacher-led instruction.

Giving Information

Variables of lesson form and quality can be divided into those that involve giving information (structuring), asking questions (soliciting), and providing feedback (reacting). The following variables apply to the function of giving information.

Structuring. Achievement is maximized when teachers not only actively present material, but structure it by beginning with overviews, advance organizers, or review of objectives; outlining the content and signaling transitions between lesson parts; calling attention to main ideas; summarizing subparts of the lesson as it proceeds; and reviewing main ideas at the end. Organizing concepts and analogies help learners link the new to the already familiar. Overviews and outlines help them to develop learning sets to use in assimilating the content as it unfolds. Rule-example-rule patterns and internal summaries tie specific information items to integrative concepts. Summary reviews integrate and reinforce the learning of major points. Taken together, these structuring elements not only facilitate memory for the information but allow for its apprehension as an integrated whole with recognition of the relationships between parts.

Redundancy/sequencing. Achievement is higher when information is presented with a degree of redundancy, particularly in the form of repeating and reviewing general rules and key concepts. The kind of redundancy that is involved in the sequential structuring built into the study by Smith and Sanders (1981) also appears important. In general, structuring, redundancy, and sequencing affect what is learned from listening to verbal presentations, even though they are not powerful determinants of learning from reading text.

Clarity. Clarity of presentation is a consistent correlate of achievement, whether measured by high inference ratings or low inference indicators such as absence of "vagueness terms" or "mazes." Knowledge about factors that detract from clarity needs to be supplemented with knowledge about positive factors that enhance clarity (for example, what kinds of analogies and examples facilitate learning, and why), but in any case, students learn more from clear presentations than from unclear ones.

Enthusiasm. Enthusiasm, usually measured by high inference ratings, appears to be related more to affective than to cognitive outcomes. Nevertheless, it often correlates with achievement, especially for older students.

Pacing/wait-time. "Pacing" usually refers to the solicitation aspects of lessons, but it can also refer to the rate of presentation of information during initial structuring. Although few studies have addressed the matter directly, data from the early grades seem to favor rapid pacing, both because this helps maintain lesson momentum (and thus minimizes inattention) and because such pacing seems to suit the basic skills learning that occurs at these grade levels. Typically, teacher presentations are short and interspersed with recitation or practice opportunities. At higher grade levels, however, where teachers make longer presentations on more abstract or complex content, it may be necessary to move at a slower pace, allowing time for each new concept to "sink in." At least, this seems to be the implication of wait-time data reported by Tobin (1980) and by Tobin and Capie (1982). Issues of pacing and wait-time during information presentation clearly need more research.

Questioning the Students

The variables in this section concern the teacher's management of public response opportunities during recitations and discussions.

Difficulty level of questions. Data on difficulty level of questions continue to yield mixed results. It seems clear that most (perhaps three-fourths) of teachers' questions should elicit correct answers, and that most of the rest should elicit overt, substantive responses (incorrect or incomplete answers) rather than failures to respond at all. Beyond these generalities, optimal question difficulty probably varies with context. Basic skills instruction requires a great deal of drill and practice, and thus frequent fast-paced drill or review lessons during which most questions are answered rapidly and correctly. However, when teaching complex cognitive content or when trying to stimulate students to generalize from, evaluate, or apply their learning, teachers will need to raise questions that few students can answer correctly (as well as questions that have no single correct answer).

Cognitive level of questions. The cognitive level of a question is conceptually separate from its difficulty level. The data reviewed here on cognitive level of question, and even meta-analyses of these and other relevant data (Winne, 1979; Redfield & Rousseau, 1981) yield inconsistent results. The data do refute the simplistic (but frequently assumed) notion that higher level questions are categorically better than lower level questions. Several studies indicate that lower level questions facilitate learning, even learning of higher level objectives. Furthermore, even when the frequency of higher level questions correlates positively with achievement, the absolute numbers on which these correlations are based typically show that only about 25% of the questions asked were classified as higher level. Thus, in general, we should expect teachers to ask more lower level than higher level questions, even when dealing with higher level content and seeking to promote higher level objectives.

These are just frequency norms, however. To develop more useful information about cognitive level of question, researchers will have to develop more complex methods of coding that take into account the teacher's goals (it seems obvious that different kinds of questions are appropriate for different goals), the quality of the questions (clarity, relevance, etc.), and their timing and appropriateness given the flow of the activity. Research on the latter issues will require shifting from the individual question to the question sequence as the unit of analysis. For example, sequences beginning with a higher level question and then proceeding through several lower level follow-up questions would be appropriate for some purposes (such as asking students to suggest a possible application of an idea, and then probing for details about how the suggested application could work). A different purpose (such as trying to call students' attention to relevant facts and then stimulate them to integrate the facts and draw an important conclusion) might require a series of lower level questions followed by a higher level question.

Clarity of question. Each teacher question should yield a (not necessarily correct) student answer. Teachers can train students to answer by showing a willingness to wait for the answer (instead of calling on someone else or giving the answer themselves). Clarity of question is also a factor; students sometimes cannot respond because questions are vague or ambiguous or because the teacher asks two or more questions without stopping to get an answer to the first one.

Post-question wait-time. Studies of science instruction have shown higher student achievement when teachers pause for about three seconds (rather than one second or less) after a question to give the students time to think before calling on one of them. This variable has not been addressed in other contexts. It seems likely, however, that length of pause following questions should vary directly with their difficulty level and especially their complexity or cognitive level. A question calling for application of abstract principles should require a longer pause than a factual question.

Selecting the respondent. Findings on this issue vary according to grade level, SES, and whole-class versus small-group setting. In the early grades, especially during small-group lessons, it is important that all students participate overtly (and roughly equally). In small-group reading lessons, this can be accomplished by using the "patterned turns" method, training the students not to call out answers or reading words, and calling on nonvolunteers as well as volunteers. In these grades, it is important to prevent assertive students from co-opting other students' response opportunities and to insure that reticent students participate regularly even though they seldom volunteer.

Student call outs usually correlate positively with achievement in low SES classes but negatively in high SES classes. This suggests the following principle: When most students are eager to respond, teachers will have to suppress their call outs and train them to respect one another's response opportunities; however, when most students are reticent, teachers will have to encourage them to participate (which may include accepting relevant call outs).

It is seldom feasible to have all students participate overtly in whole class lessons, let alone to insure that all participate equally. This need not present a problem even in the lower grades in subjects such as spelling or arithmetic computation, where practice and assessment can be accomplished through written exercises. It may present a dilemma, however, for primary grade teachers working on objectives that call for overt verbal practice and for teachers at any level who want to make assignments that call for students to make verbal presentations to the group (speeches, research reports). Here, it may be necessary to divide the class into groups or to schedule only a few presentations per day and use the rest of the period for faster paced activities.

Except as noted in the previous paragraph, overt verbal participation in lessons does not seem to be an important achievement correlate in the upper grades. Still, rather than interact with the same few students most of the time, teachers in these grades probably should encourage volunteering (pausing after asking questions, to give students time to think and raise their hands, will help here) and call on nonvolunteers frequently (especially when they are likely to be able to respond correctly).

Waiting for the student to respond. Once teachers do call on students (especially nonvolunteers), usually they should wait until the students offer a substantive response, ask for help or clarification, or overtly say "I don't know." Sometimes, however, especially in whole-class lessons where lengthy pauses threaten continuity or momentum, it will be necessary for the teacher to curtail the pause by making one of the reacting moves discussed in the following section.

Reacting to Student Responses

Once the teacher has asked a question and called on a student to answer, the teacher then must monitor the student's response (or lack of it) and react to it.

Reactions to correct responses. Correct responses should be acknowledged as such, because even if the respondent knows that the answer is correct, some of the onlookers may not. Ordinarily (perhaps 90% of the time) this acknowledgement should take the form of overt feedback, which may range from brief head nods through short affirmation statements ("right," "yes") or repetition of the answer, to more extensive praise or elaboration of the answer. Such overt affirmation can be omitted on occasion, such as during fast-paced drills in which the students understand that the teacher will simply move on to the next question if the previous question is answered correctly.

Although it is important for teachers to give feedback so that everyone knows that an answer was correct, it usually is not important to praise the student who supplied the answer. Such praise is often intrusive and distracting; it may even embarrass the recipient, especially if the accomplishment was not especially praiseworthy in the first place. In any case, teachers who maximize achievement are sparing rather than effusive in praising correct answers. To the extent that such praise is effective, it is more likely to be effective when it is specific rather than global and when it is used with low SES or dependent/anxious students rather than with high SES or assertive/confident students.

Reacting to partly correct responses. Following responses that are incomplete or only partly correct, teachers ordinarily should affirm the correct part and then follow up by giving clues or rephrasing the question. If this does not succeed, the teacher can give the answer or call on another student.

Reacting to incorrect responses. Following incorrect answers, teachers should begin by indicating that the response is not correct. Almost all (99%) of the time, this negative feedback should be simple negation rather than personal criticism, although criticism may be appropriate for students who have been persistently inattentive or unprepared.

After indicating that the answer was incorrect, teachers usually should try to elicit an improved response by rephrasing the question or giving clues. Such response improvement attempts are likely to be facilitative when they are generally successful, but teachers should avoid pointless pumping in situations where questions cannot be broken down or the student is too confused or anxious to profit from further questioning.

Sometimes the feedback following an incorrect answer should include not only the correct answer but a more extended explanation of why the answer is correct or how it can be determined from the information given. Such extended explanation should be included in the feedback whenever the respondent (or others in the class) might not "get the point" from hearing the answer alone, as well as at times when a review or summary of part of the lesson is needed.

Reacting to no response. Teachers should train their students to respond overtly to questions, even if only to say, "I don't know." If waiting has not produced an overt response, teachers should probe ("Do you know?"), elicit an overt response, and then follow up by giving feedback, supplying the answer, or calling on someone else (depending on the student's response to the probe).

Reacting to student questions and comments. Teachers should answer relevant student questions or redirect them to the class and incorporate relevant student comments into the lesson. Such use of student ideas appears to become more important with each succeeding grade level, as students become both more able to contribute useful ideas and more sensitive to whether teachers treat their ideas with interest and respect.

Handling Seatwork and Homework Assignments

Although independent seatwork is probably overused and is not a substitute for active teacher instruction or for drill/recitation/discussion opportunities, seatwork (and homework) assignments provide needed practice and application opportunities. Ideally, such assignments will be varied and interesting enough to motivate student engagement, new or challenging enough to constitute meaningful learning experiences rather than pointless busywork, and yet easy enough to allow success with reasonable effort. For assignments on which students are expected to work on their own, success rates will have to be very high--near 100%. Lower (although still generally high) success rates can be tolerated when students who need help can get it quickly.

Student success rates, and the effectiveness of seatwork assignments generally, are enhanced when teachers explain the work and go over practice examples with the students before releasing them to work independently. Furthermore, once the students are released to work independently, the work goes more smoothly if the teacher (or an aide) circulates to monitor progress and provide help when needed. If the work has been well chosen and well explained, most of these helping interactions will be brief, and, at any given time, most students will be progressing smoothly through the assignment rather than waiting for help.

Students should know what work they are accountable for, how to get help when they need it, and what to do when they finish. Performance should be monitored for completion and accuracy, and students should receive timely and specific feedback. When the whole class or group has the same assignment, review of the assignment can be part of the next day's lesson. Other assignments will require more individualized feedback. Where performance is poor, teachers should provide not only feedback but reteaching and follow up assignments designed to insure that the material is mastered.

Among responses to seatwork and homework performance, feedback and follow up are more closely related to achievement than praise or reward. Even so, positive relationships have been reported for praise, symbolic rewards, and token reinforcement (at least in the early grades). Such rewards may facilitate learning if tied to complete and correct performance on assignments.

Context-Specific Findings

Even the most widely replicated process-product relationships usually must be qualified by references to the context of instruction. Usually, these interactions with context involve minor elaborations of main trends, but occasionally, as in the Brophy and Evertson (1976) or the Solomon and Kendall (1979) studies, interactions are more powerful than main effects and suggest qualitatively different treatment for different groups of students. Certain interaction effects appear repeatedly and constitute well established findings.

Grade level. In the early grades, classroom management involves a great deal of instruction in desired routines and procedures. Less of this instruction is necessary in the later grades, but it becomes especially important to be clear about expectations and to follow up on accountability demands. Lessons in the early grades involve basic skills instruction, often in small groups, and it is important that each student participate overtly and often. In later grades, lessons typically are with the whole class and involve applications of basic skills or consideration of more abstract content. Overt participation is less important than factors such as teachers' structuring of the content, clarity of statements and questions, and enthusiasm. The praise and symbolic rewards that are common in the early grades give way to the more impersonal and academically centered instruction common in the later grades, although it is important for teachers in the later grades to treat students' contributions with interest and respect.

Student SES/ability/affect. SES is a proxy for a complex of correlated cognitive and affective differences between subgroups of students. The cognitive differences involve IQ, ability, or achievement levels. Interactions between process-product findings and student SES or achievement level indicate that low SES/low achieving students need more control and structuring from their teachers (more active instruction and feedback, more redundancy, and smaller steps with higher success rates). This will mean more review, drill, and practice, and thus more lower level questions. Across the school year, it will mean exposure to less material, but with emphasis on mastery of the material that is taught and on moving students through the curriculum as briskly as they are able to progress.

Affective correlates of SES include the degree to which students feel secure and confident versus anxious or alienated in the classroom. High SES students are more likely to be confident, eager to participate, and responsive to challenge. They want respect and require feedback, but usually do not require a great deal of encouragement or praise. They thrive in an atmosphere that is academically stimulating and somewhat demanding. Low SES students are more likely to require warmth and support in addition to good instruction, and to need more encouragement for their efforts and praise for their successes. It is especially important to teach them to respond overtly rather than remain passive when asked a question, and to be accepting of their (relevant) call outs and other academic initiations when they do occur.

Teacher's intentions/objectives. What constitutes appropriate instructional behavior will vary with the teacher's objectives. This factor has rarely been studied directly, but relevant principles can be inferred easily from the data reviewed. First, as an extension of the principle of student opportunity to learn, it seems obvious that instruction designed to achieve particular objectives should include teacher presentation of, and student opportunity to practice or apply, content relevant to those objectives. This, in turn, has implications about what methods are appropriate. To the extent that the students need new information, they are likely to need group lessons featuring a presentation in which the teacher supplies information followed by recitation or discussion opportunities. The appropriateness of follow up practice or application opportunities would depend on the objectives. When it is sufficient that the students be able to reproduce knowledge on cue, routine seatwork assignments and tests might suffice. However, if students are expected to integrate broad patterns of learning or apply them to their everyday lives, it will be necessary to schedule activities that involve problem solving, decision making, essay composition, preparation of research reports, or construction of some product. In general, the nature and cognitive level of the information given and the questions asked during an activity should depend on the objectives being pursued and the place of the activity within the anticipated progression through the curriculum.

Other. Some findings are specific to particular contexts. For example, the principles put forth by Anderson, Evertson, and Brophy (1982) are specific to small group instruction in the primary grades, and several studies included variables that are specific to subject matter (such as concentration on word attack versus comprehension in reading instruction). These and other context factors must be considered in attempting to generalize from any study.

Power and Limits of the Data

The last 15 years have finally produced an orderly knowledge base linking teacher behavior to achievement. Although just a beginning, this is a major advance over what was available previously. If applied with proper attention to its limits, this knowledge base should help improve teacher education and teaching practice. Several important limits and qualifications need to be kept in mind, however.

One is that the causal relationships that explain linkages between teacher behavior and student achievement are not always clear, and even when they are, process-product relationships do not translate directly into prescriptions for teaching practice. In the case of correlations between teacher behaviors and achievement, positive correlations do not necessarily indicate that the teacher behavior should be maximized (even within the observed range, let alone the theoretical range). Thus it would be inappropriate to conclude that teachers should always wait at least three seconds for a response to a question, should never criticize students, or should never schedule independent seatwork.

To develop sensible recommendations about teacher behaviors, one must consider their means and ranges of variation. A positive correlation for a behavior that happens regularly must be interpreted differently from a positive correlation for a behavior that occurs only rarely. In addition, one must consider the contexts within which the behavior occurs and its patterns of relationship with other teacher behaviors and with student behaviors. In what context is this teacher behavior an option? What other options are available in the same contexts? When is this behavior the option of choice, and why? Answering such questions requires knowledge about process-process as well as process-product relationships (and more generally, a familiarity with classrooms and how they work).

In effect, then, although it is necessary to study teacher behaviors individually in order to establish their specific relationships to achievement, and although it is necessary to strip away much of the context in which these behaviors are embedded in order to accumulate a large enough sample of comparable behaviors to allow the use of inferential statistics, interpretation of the relationships thus identified requires reconsideration of the teacher behaviors as parts of larger patterns occurring in particular contexts. In trying to develop guidelines about when and for how long teachers should wait for students to answer a question, one must consider such factors as the nature of the question, whether the student seems to be thinking about the question or is likely to profit from additional time, and whether further waiting might endanger the lesson's continuity or momentum.

Different patterns might be functionally equivalent. For example, it may make no important difference whether the three main points of a presentation are summarized at the beginning or the end of the presentation (so long as they are summarized) or whether a mathematics computation review is done with flash cards during a lesson or through a seatwork assignment afterwards. Functionally equivalent patterns such as this have rarely been considered, let alone investigated systematically (see Good & Power, 1976, for discussion of functionally equivalent classroom experiences).

The fact that there may be different but functionally equivalent paths to the same outcome is but one reason why data linking teacher behavior to achievement should not be used for teacher evaluation or accountability purposes. If teachers are to be evaluated according to the achievement they produce, then this achievement should be measured directly. Information on short-term outcomes such as academic learning time or performance on assignments might be of some use, but it would be inappropriate to penalize teachers for failing to follow overly rigid behavioral prescriptions if they produced as much achievement as the teachers who did follow the prescriptions.

Another reason why the data presented here cannot be used in any simple fashion for evaluating teachers is that achievement gain was the only outcome considered in detail. Teachers vary not only in their success in producing achievement, but in their success in fostering positive attitudes, personal development, and good group relations. Unfortunately, success on one of these dimensions does not necessarily imply success on the others. It is possible to optimize progress along several dimensions simultaneously to some degree, but beyond some point, further progress toward one objective will come at the expense of progress toward others. Even ideal teaching will involve trade-offs rather than optimizing in an absolute sense (Clark, 1982; Evertson, 1979; Peterson, 1979; Schofield, 1981).

Another limit to these data is that the correlational findings were based on natural variation in existing classroom practices, and most of the experiments involved practices previously observed occurring spontaneously. Several implications follow. One is that generalization of these data is probably limited to traditionally taught classrooms (they would not apply to totally individualized approaches, for example). Another is that prescriptions for application probably should remain within the ranges of teacher behavior observed in these studies. Simpleminded extrapolations beyond those ranges (such as, if 15 minutes of homework per night is good, two hours per night would be eight times better) are not supported by the data and probably are counterproductive. A third point to consider is that naturalistic data reflect the practices prevalent in the time and place in which they were collected (primarily the United States in the 1970s, in this case). Compared to schools in Europe and Japan, American schools in the 1970s probably featured less active (whole-class or small-group) instruction by teachers, less content coverage per unit of time, and less time on task (c.f. Dalton & Willcocks, 1983). Consequently, quantity of instruction and opportunity to learn factors were among the strongest correlates of achievement. In other countries, however, or wherever content coverage is uniformly high (and variance is low), qualitative measures of teaching might correlate more strongly with achievement than quantitative measures.

Finally, most findings must be qualified by grade level, type of objective, type of student, and other context factors. This creates dilemmas for teachers working with heterogeneous classes. Furthermore, even within context, it seems likely that all relationships are ultimately curvilinear. Too much of even a generally good thing is still too much.

At least two common themes cut across the findings, despite the need for limitations and qualifications. One is that academic learning is influenced by the amount of time that students spend engaged in appropriate academic tasks. The second is that students learn more efficiently when their teachers first structure new information for them and help them relate it to what they already know, then monitor their performance and provide corrective feedback during recitation, drill, practice, or application activities. For a time, these generalizations seemed confined to the early grades or to basic rather than more advanced skills. However, it now appears that they apply to any body of knowledge or set of skills that has been sufficiently well organized and analyzed so that it can be presented (explained, modeled) systematically and then practiced or applied during activities that call for student performance that can be evaluated for quality and (where incorrect or imperfect) given corrective feedback.

This certainly includes aspects of reading comprehension and mathematics problem solving in addition to word attack and mathematics computation, and it probably includes aspects of complex learning that are not usually thought of as attainable through systematic teaching (developing learning-to-learn skills, creative writing, artistic expression). Even for higher level, complex learning objectives, guidance through planned sequences of experience is likely to be more effective than unsystematic trial and error.

It should be noted that the instruction involved in such higher level activities is often highly complex and demanding. Instead of supplying simple algorithms to be imitated or giving correct answers to factual questions, effective instructors working at higher levels must be able to develop apt analogies or examples that will enable students to relate the new to the familiar or the abstract to the concrete, identify key concepts that help to organize complex bodies of information, model problem solving processes that involve judgment and decision making under conditions of uncertainty, and diagnose and correct subtle misconceptions in students' thinking. These are complex, demanding, and yet essential activities; they should neither be demeaned as intrusive "teacher talk" nor confused with the relatively simple "telling" or giving of "right answers" that occur in basic skills lessons in the early grades.

Finally, it should be stressed that there are no shortcuts to successful attainment of higher level learning objectives. Such success will not be achieved with relative ease through discovery learning by the students. Instead, it will require considerable instruction from the teacher and thorough mastery of basic knowledge and skills that must be integrated and applied in the process of "higher level" performance. Development of basic knowledge and skills to the necessary levels of automatic and errorless performance will require a great deal of drill and practice. Thus, drill and practice activities should not be slighted as "low level." They appear to be just as essential to complex and creative intellectual performances as they are to the performance of a virtuoso violinist.

Methodological Notes

Methodological issues were discussed in detail in an earlier section, and pertinent aspects of methodology were mentioned in discussing individual studies. Rather than repeat all of that here, we will merely call attention to a few salient points. Most of these concern the need for more targeted and refined measurement. Data compiled by forcing unselected interaction into a few general categories and then computing frequencies are not very useful. Better data will result from thought and planning devoted to issues such as the following.

What are the teacher and student behaviors of interest, and in what contexts do they occur? If they occur only in certain contexts, data collection must be planned for these contexts, and other contexts can be ignored. However, within the range of relevant contexts, behavior may vary in meaning or in patterns of correlation with other variables. Thus, the behavior may have to be measured somewhat differently in different contexts. In any case, it will be important to record the data so that tallies can be analyzed separately for each context in addition to being combined across contexts.

Within context, when is the behavior possible or likely to occur? If this question can be answered clearly, it will be possible to supplement frequency scores indicating the rate of occurrence of the behavior with percentage scores (for example, percentage of lessons begun with an overview) reflecting the relative frequency of the behavior in situations in which it could be expected. These two types of scores carry different information.

Does the behavior usually occur as part of a predictable sequence? If so, the coding should be planned to allow examination of entire sequences in addition to separate examination of the component behaviors. Patterns of initiation and reaction should be retained in the coding system so that proactive teacher behaviors can be separated from reactive teacher behaviors that occur in response to student initiations.

Are there important distinctions concerning the quality, timing, or appropriateness of the behavior that can be built into the coding system? Even the most sophisticated contemporary coding systems use relatively crude, global category definitions that could be improved considerably through differentiation of qualitatively different subtypes or coding the appropriateness of behaviors in addition to merely noting their occurrences. For example, Brophy (1981) reviewed a wide range of literature on teacher praise, noting that such praise has different purposes and meanings in different contexts. He concluded that the quality of teacher praise is more important than its frequency and offered the guidelines shown in Table 4. Research that built some of these qualitative distinctions into the measurement of teacher praise would probably produce more orderly and meaningful results than the research done to date using cruder measures. Note that praise is just an example; conceptualization and measurement of most of the other teacher behavior variables discussed in this chapter are equally crude and in need of elaboration.

Existing findings on quantity of instruction are stronger and more consistent than the findings on quality, because so many findings were derived from naturalistic situations where teachers varied drastically in their allocation of time to academic activities and in their classroom organization and management skills. The differences in student opportunity to learn created by these differences in time allocation and classroom management probably overwhelmed, and thus masked, the effects of whatever differences in quality of instruction occurred. To study quality differences, it will be necessary to control quantity differences, at minimum by restricting samples to teachers who are skilled in classroom management and similar in their goals and time allocations, and sometimes even by scripting or otherwise controlling the amount and nature of instruction.

Table 4

Guidelines for Effective Praise

Effective Praise	Ineffective Praise
1. is delivered contingently	1. is delivered randomly or unsystematically
2. specifies the particulars of the accomplishment	2. is restricted to global positive reactions
3. shows spontaneity, variety, and other signs of credibility; suggests clear attention to the student's accomplishment	3. shows a bland uniformity that suggests a conditioned response made with minimal attention
4. rewards attainment of specified performance criteria (which can include effort criteria, however)	4. rewards mere participation, without consideration of performance processes or outcomes
5. provides information to students about their competence and the value of their accomplishments	5. provides no information at all or gives students information about their status
6. orients students toward better appreciation of their own task-related behavior and thinking about problem solving	6. orients students toward comparing themselves with others and thinking about competing
7. uses students' own prior accomplishments as the context for describing present accomplishments	7. uses the accomplishments of peers as the context for describing students' present accomplishments
8. is given in recognition of noteworthy effort or success at difficult (for *this* student) tasks	8. is given without regard to the effort expended or the meaning of the accomplishment
9. attributes success to effort and ability, implying that similar successes can be expected in the future	9. attributes success to ability alone or to external factors such as luck or (easy) task difficulty
10. fosters endogenous attributions (students believe that they expend effort on the task because they enjoy the task and/or want to develop task-relevant skills)	10. fosters exogenous attributions (students believe that they expend effort on the task for external reasons--to please the teacher, win a competition or reward, etc.)

Table 4 (continued)

Guidelines for Effective Praise

Effective Praise	Ineffective Praise
11. focuses students' attention on their own task-relevant behavior	11. focuses students' attention on the teacher as an external authority figure who is manipulating them
12. fosters appreciation of, and desirable attributions about, task relevant behavior after the process is completed	12. intrudes into the ongoing process, distracting attention from task-relevant behavior

Note. Table 5 is from Brophy (1981, Spring) Teacher praise: A functional analysis. Review of Educational Research, p. 5-32. Copyright 1981, American Educational Research Association, Washington, D.C.

How can the teacher behavior be sampled and measured reliably? If the study will rely on naturalistic observation, it will be important to observe in contexts in which the behavior appears frequently and to observe often and long enough to build up a reliable sample of behavior.

Is there congruence between the content taught, the categories in the observation system, and the content tested? The content taught in the different classes to be observed should be identical (otherwise, differences in curricula will be confounded with differences in methods), and the test should be a valid and reliable sampling of that content. Where relevant, the test data should allow for separate analysis of different types or levels of learning as well as distinctions such as whether the material was specifically taught by the teacher or merely included in the text, or whether the items tapped intentional or incidental learning. In addition, the coding categories should reflect the content taught--the categories used for coding small-group reading instruction in 1st grade should be very different from those used for coding whole-class science instruction in 12th grade.

How should the data be reported? At minimum, both descriptive information (means, variance) and process-product information (correlation or regression coefficients) should be provided for each separate classroom process variable. Results of multiple regression analyses and data for combination scores can be given as well, but in addition to, rather than instead of, basic descriptive and correlational data for each variable. Relevant context information should be supplied and mentioned as qualifiers on potential generalization of the findings, and any suggested prescriptions for teaching practice should place the teaching behaviors back into context and take into account the naturally occurring limits and variance within which the obtained correlational relationships occurred.

Next Steps in Research on Teacher Effects

To enhance its value for both theory and practice, research on teacher effects needs not only methodological improvements but expansion into new areas. For example, most existing research yields conclusions about lessons or lesson components, but little information is available on larger units of instruction. What are the characteristics of an effective week or unit? How are concepts learned in one unit used effectively in the next? How can units be designed to allow for distributed practice, meaningful integration of learning, and transfer or application?

A related point is that more attention needs to be given to consecutive sequences of instruction. How does information gathered in the process of interacting with students today affect the teacher's instructional behavior or assignments tomorrow? What changes should occur in the nature and length of lesson components (presentation of new information, recitation, drill, etc.) as teachers initiate and move through a unit? To study these issues of instructional redundancy, integration of concepts, and teachers' processing and use of information gathered during teaching, researchers will have to focus on the instructional unit rather than the lesson as their unit of analysis and to observe over several consecutive days rather than spread observations across the term.

More thick description and microanalysis of how lessons and lesson components are accomplished by teachers are also needed. For example, Good and Grouws (1979a,1979b) urged teachers to include in their lessons a development phase in which they would present concepts, give examples, demonstrate through modeling, and the like. They gave guidelines about how much time to spend in developmental phases of lessons, but not much qualitative advice, let alone step-by-step instructions, about how to accomplish development segments. One logical next step for their research would be to concentrate attention on development portions of lessons in order to become more prescriptive about the nature and sequencing of steps to include, about the effectiveness of different kinds of examples, and so on (see Good et al., 1983 for related discussion).

Attention should be payed to teachers' goals and intentions. Researchers need to know what teachers are trying to accomplish in order to interpret, and make useful contextual and qualitative distinctions for coding, their behaviors. Depending on the teacher's intention at the time, behavior such as asking a particular question or praising a particular student's response may or may not be appropriate.

More attention needs to be given to higher level instruction (higher in terms of both grade level and cognitive level). It will be especially important to control the objectives that teachers are working toward in this context. Presently, debates on "cognitive level of instruction" seem to resolve to conflicts about curriculum (what should be taught) rather than method (how it should be taught). Progress toward resolution is unlikely until this confusion is eliminated and appropriate research is conducted. Curriculum issues should be addressed by minimizing the variation in teaching methods or at least the variance in outcomes that can be attributed to differences in teaching methods (ideally, by insuring that each curriculum is taught as well as it can be taught). Method issues must be addressed by holding curriculum constant. Productive research on teaching to higher level objectives may require not only controlling the content taught in a general sense, but scripting teachers' behavior during lessons and controlling the curriculum materials and assignments to which the students are exposed.

Many questions about effective instruction have either not yet been studied or not studied appropriately: the selection and sequencing of questions to include in recitations or discussions designed to achieve particular goals; the nature of the diagnosis and feedback that should occur when teachers monitor progress while students work on assignments; the relative advantages of various accountability procedures, scoring and grading practices, and review and reteaching practices that are applied to completed seatwork assignments; qualitative aspects of teacher presentations other than clarity (usefulness of examples and analogies, density and sequencing of information, length of information presentation segments, and placement with respect to questioning or practice segments); the relative advantages of various questions, tasks, and assignments that students are to work on independently; and questions about how effective instruction might evolve during the course of a unit or a school year.

Integrating Teacher Effects Research with Other Research

Instruction and its relationships to achievement can be isolated for purposes of analysis, but in reality instruction always occurs within particular contexts. Consequently, in designing and speculating about the implications of research on teacher effects, it is useful to consider such research in conjunction with research on factors other than teacher behavior and student achievement. The following three types of research seem especially apropos.

Subject matter instruction. Research on instruction in topics within specific subject areas supplements the process-outcome findings reported here. Research in reading instruction, for example, has shown that concepts can be taught more effectively using certain types and sequences of examples rather than other types or sequences (Engelmann & Carnine, 1982), and that students can be taught not only factual knowledge and word attack skills, but also the higher level concepts and learning-to-learn skills needed for effective reading comprehension (Adams, Carnine, & Gersten, 1982; Brown, Campione, & Day, 1981).

Research in mathematics and science instruction has shown that many concepts are counterintuitive or otherwise difficult to grasp and retain, not only for students but also for teachers and other adults. Consequently, teachers with limited backgrounds in certain subject matter areas may teach incorrect content or fail to recognize and correct their students' distorted understandings (c.f. Eaton, Anderson, & Smith, 1984). Clearly, the effectiveness of lessons will vary with teachers' interest in and knowledge about the content being taught.

More generally, in delineating the contexts within which instruction occurs, researchers need to pay more specific attention not only to types of lessons and lesson components (Brophy, 1979; Berliner, 1983), but also to the scope and sequence of the curriculum and to the specific subject matter goals and content taught in particular lessons (see Romberg, 1983, for examples in the area of mathematics instruction).

Student mediation of instruction. Teachers' instructional objectives are mediated not only by teacher behavior but by academic tasks that teachers present to students (Doyle, 1983) and by students' individualized responses to instruction and academic activities. Students will carry different meanings away from the same lecture or demonstration (Winne & Marx, 1982), respond differentially to teacher behaviors such as praise (Brophy, 1981; Morine-Dershimer, 1982; Weinstein, 1983), and demonstrate diverse needs for structure or autonomy (Ebmeier & Good, 1979; Janicki & Peterson, 1981). Students can profitably teach one another or work together under certain conditions

(Slavin, 1980), although some grouping arrangements work better than others (Webb, 1980).

The effects of many teacher behaviors that appear to facilitate achievement (clear presentations, appropriate difficulty level of questions, structuring of the content, specificity of feedback and praise) are mediated via students' immediate information processing. These teacher behaviors provide students with information or engage them with content so that the new information is assimilable, short term memory is not overloaded (which helps make assimilation possible), connections are made between existing knowledge and the new information, and "chunking" and other efficient mechanisms for processing and retaining information are developed through engagement in appropriate practice and application activities. Discussion of these short term cognitive outcomes of instruction (changes in students' concepts or content-related information processing abilities) may be more useful to teachers or teacher educators than discussion of scores on norm-referenced achievement tests. More generally, better articulation of research on teaching with research on students' mediation of classroom events should help researchers to understand the causal linkages underlying process-outcome relationships and discover unintended side effects of teacher behaviors. The eventual result should be a grounded theory of teaching and its effects.

Other outcome variables. Research on achievement outcomes needs to be articulated with research on other student outcomes. Effects on student attitudes toward the teacher, the subject matter, or the class were reported in some of the studies reviewed here. Student attitudes were linked most closely to measures of teacher warmth and student orientation: praise, use of student ideas, willingness to listen to students and respect their contributions, and socializing with students in addition to instructing them. These teacher behaviors are mostly just different from, rather than either similar or contradictory to, the teacher behaviors associated with achievement, and the two sets of behaviors are compatible to some extent. However, students are likely to have more positive attitudes toward moderately demanding teachers than toward highly demanding teachers.

Few researchers of teacher effects on achievement have gathered information on outcomes such as the development of independence, good work habits, social skills, or personal adjustment and mental health. Nor have studies concerned with these outcomes typically measured achievement. Clearly, more research that addresses these multiple outcomes simultaneously within the same study is needed to develop information about what trade-offs must be faced and about what can realistically be accomplished in typical classrooms.

Certain instructional methods have predictable effects on outcomes other than achievement. For example, achievement differences among students are more salient in classes that are subdivided into ability groups or taught routinely as whole classes involving public performance and evaluation than they are in classes that feature individualized instruction or flexible small-group assignments based on factors other than achievement (Bossert, 1979; Rosenholtz & Wilson, 1980; Weinstein, 1983). Social relationships are likely to focus on peers of similar ability level in the former classes, but to involve a broader range of peers in the latter classes. It is also true that the social aspects of education may have important effects on what is learned or how well (Florio, 1978; Eder, 1981; Rosenholtz & Cohen, 1983). Eventually, it will be necessary to integrate research on teacher effects with research on classroom composition (e.g., Bossert, 1979), classroom ecology (Doyle, 1979; Hamilton, 1983), and student perceptions (Weinstein, 1983) to develop a more complete picture of how schooling influences student outcomes.

Conclusion

Comparison of this paper with related chapters in the first and second editions of the *Handbook of Research on Teaching* show how much more is now known about teacher effects on achievement than in 1963 or even 1973. The myth that teachers do not make a difference in student learning has been refuted, and programmatic research reflecting the description-correlation-experimentation loop called for by Rosenshine and Furst (1973) has begun to appear. As a result, the fund of available information on producing student achievement (especially the literature relating to the general area of classroom management and to the subject areas of elementary reading and mathematics instruction) has progressed from a collection of disappointing and inconsistent findings to a small but well established knowledge base that includes several successful field experiments.

Although illustrating that instructional processes make a difference, this research also shows that complex instructional problems cannot be solved with simple prescriptions. In the past, when detailed information describing classroom processes and linking them to outcomes did not exist, educational change efforts were typically based on simple theoretical models and associated rhetoric calling for "solutions" that were both oversimplified and overly rigid. The data reviewed here should make it clear that no such solution can be effective because what constitutes effective instruction (even if attention is restricted to achievement as the sole outcome of interest) varies with context. What appears to be just the right amount of teacher academic demand (or structuring of content, or praise, etc.) for one class might be too much for a second class but not enough for a third class. Even within the same class, what constitutes effective instruction will vary according to subject matter, group size, and the specific instructional objectives being pursued.

Elitist critics often undervalue teaching or even suggest that anyone can teach ("Those who can, do; those who can't, teach"). The data reviewed here refute this myth as well. Although it may be true that most adults could survive in the classroom, it is not true that most could teach effectively. Even trained and experienced teachers vary widely in how they organize the classroom and present instruction. Specifically, they differ in the expectations and achievement objectives they hold for themselves, their classes, and individual students; how they select and design academic tasks; and how actively they instruct and communicate with students about academic tasks. Those who do these things successfully produce significantly more achievement than those who do not, but doing them successfully demands a blend of energy, motivation, and communication and decision-making skills that many teachers, let alone ordinary adults, do not possess.

Improvement of education must begin with recruitment of capable teachers, followed by retention of those teachers in the teacher role. Preservice and inservice teacher education in both subject matter and pedagogy are also essential, however. This includes familiarizing teachers with the findings reviewed here. This may sound gratuitous, but many teachers, even recently trained ones, are not aware of important concepts and findings from research on teaching.

It is important that this information be presented in ways that respect the uniqueness of each classroom and recognize that classrooms are complex social settings in which teachers must process a great deal of information rapidly, deal with several agendas simultaneously, and make quick decisions throughout the day. Thus, rather than trying to translate it into overly rigid or generalized prescriptions, teacher educators should present this information to teachers within a decision-making format that enables them to examine concepts critically and adapt them to the particular contexts within which they teach (for an illustration of this, see Amarel, 1981). Research on how teacher education programs can accomplish this effectively is badly needed.

Also needed, of course, is more research on teaching in general and on teacher effects in particular. Despite the successes of the 1960s and 1970s, progress has slowed noticeably of late. In part, this is because the field is in a natural period of consolidation following a period of rapid development of new findings using newly developed techniques. However, reduction in support for research on teaching has been another factor. Just as a knowledge base about teaching and teacher education was finally becoming established, the budget for the National Institute of Education was being decimated repeatedly. Adjusted for inflation, federal support of educational research is now below one-third of what it was even a few years ago. We hope that this trend will be reversed, so that authors writing about teacher effects research in the next handbook will also be able to report the kind of progress that we have been able to report in this one.

References

Acland, H. (1976). Stability of teacher effectiveness: A replication. Journal of Educational Research, 69, 289-292.

Adams, A., Carnine, D., & Gersten, R. (1982). Instructional strategies for studying content area texts in the intermediate grades. Reading Research Quarterly, 18, 27-55.

Alexander, L., Frankiewicz, R., & Williams, R. (1979). Facilitation of learning and retention of oral instruction using advance and post organizers. Journal of Educational Psychology, 71, 701-707.

Amarel, M. (1981, April). Literacy: The personal dimension. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, CA.

Anderson, L., Evertson, C., & Brophy, J. (1979). An experimental study of effective teaching in first-grade reading groups. Elementary School Journal, 79, 193-223.

Anderson, L., Evertson, C., & Brophy, J. (1982). Principles of small-group instruction in elementary reading (Occasional Paper No. 58). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University.

Anderson, O. (1969). Structure in teaching: Theory and analysis. New York: Teachers College Press.

Arehart, J. (1979). Student opportunity to learn related to student achievement of objectives in a probability unit. Journal of Educational Research, 72, 253-269.

Armento, B. (1977). Teacher behaviors related to student achievement on a social science concept test. Journal of Teacher Education, 28, 46-52.

Aspy, D. (1969). The effect of teacher-offered conditions of empathy, positive regard, and congruence upon student achievement. Florida Journal of Educational Research, 11, 39-48.

Aspy, D. (1972). Toward a technology for humanizing education. Champaign, IL: Research Press.

Ausubel, D. (1968). Educational psychology: A cognitive view. New York: Holt, Rinehart, & Winston.

Barr, R., & Dreeben, R. (1978). Instruction in classrooms. In L. Shulman (Ed.), Review of research in education (Vol. 5). Itasca, Illinois: Peacock.

Bellack, A., Hyman, R., Smith, F., & Kliebard, H. (1966). The language of the classroom. New York: Columbia University Press.

Bennett, N. (1976). Teaching styles and pupil progress. London: Open Books Publishing Limited.

Bennett, N., Desforges, C., Cockburn, A., & Wilkinson, B. (1981). The quality of pupil learning experiences: Interim report. Lancaster, England: Centre for Educational Research and Development, University of Lancaster.

Berliner, D. (1976). Impediments to the study of teacher effectiveness. Journal of Teacher Education, 27, 5-13.

Berliner, D. (1977). Impediments to measuring teacher effectiveness. In G. Borich & K. Fenton (Eds.), The appraisal of teaching: Concepts and process. Reading, MA: Addison-Wesley.

Berliner, D. (1979). Tempus educare. In P. Peterson & H. Walberg (Eds.). Research on teaching: Concepts, findings, and implications. Berkeley, CA: McCutchan.

Berliner, D. (1983). Developing conceptions of classroom environments: Some light on the T in classroom studies of ATI. Educational Psychologist, 18, 1-13.

Berliner, D., Fisher, C., Filby, N., & Marliave, R. (1978). Executive Summary of Beginning Teacher Evaluation Study. San Francisco, CA: Far West Regional Laboratory for Educational Research and Development.

Berliner, D., & Tikunoff, W. (1976). The California Beginning Teacher Evaluation Study: Overview of the ethnographic study. Journal of Teacher Education, 27(1), 24-30.

Berliner, D., & Tikunoff, W. (1977). Ethnography in the classroom. In G. Borich & K. Fenton (Eds.), The appraisal of teaching: Concepts and process. Reading, MA: Addison-Wesley.

Bettencourt, E., Gillett, M., Gall, M., & Hull, R. (1983). Effects of teacher enthusiasm training on student on-task behavior and achievement. American Educational Research Journal, 20, 435-450.

Blaney, R. (1983). Effects of teacher structuring and reacting on student achievement. Elementary School Journal, 83, 569-577.

Blank, M. (1973). Teaching learning in the preschool: A dialogue approach. Columbus, OH: Merrill.

Bloom, B., Engelhart, M., Furst, E., Hill, W., & Krathwohl, D. (Eds.). (1956). Taxonomy of educational objectives: Handbook I: Cognitive domain. New York: David McKay.

Boak, R., & Conklin, R. (1975). The effect of teachers' levels of interpersonal skills on junior high school students' achievement and anxiety. American Educational Research Journal, 12, 537-549.

Borg, W. (1979). Teacher coverage of academic content and pupil achievement. Journal of Educational Psychology, 71, 635-645.

Borg, W. (1980). Time and school learning. In C. Denham & A. Lieberman (Eds.), Time to learn. Washington, D.C.: National Institute of Education, U.S. Department of Education.

Borich, G., & Fenton, K. (Eds.). (1977). The appraisal of teaching: Concepts and process. Reading, MA: Addison-Wesley.

Bossert, S. (1979). Task and social relationships in classrooms: A study of classroom organization and its consequences. In American Sociological Association, Arnold and Caroline Rose Monograph Series. New York: Cambridge University Press.

Brophy, J. (1973). Stability of teacher effectiveness. American Educational Research Journal, 10, 245-252.

Brophy, J. (1979). Teacher behavior and its effects. Journal of Educational Psychology, 71, 733-750.

Brophy, J. (1981). Teacher praise: A functional analysis. Review of Educational Research, 51, 5-32.

Brophy, J., & Evertson, C. (1974a). Process-product correlations in the Texas Teacher Effectiveness Study: Final Report (Research Report 74-4). Austin: Research and Development Center for Teacher Education, University of Texas at Austin. (ERIC ED 091 094)

Brophy, J., & Evertson, C. (1974b). The Texas Teacher Effectiveness Project: Presentation of non-linear relationships and summary discussion (Research Report No. 74-6). Austin: Research and Development Center for Teacher Education, University of Texas at Austin. (ERIC ED 099 345)

Brophy, J., & Evertson, C. (1976). Learning from teaching: A developmental perspective. Boston, MA: Allyn and Bacon.

Brophy, J., & Evertson, C. (1978). Context variables in teaching. Educational Psychologist, 12, 310-316.

Brown, A., Campione, J., & Day, J. (1981). Learning to learn: On training students to learn from texts. Educational Researcher, 2, 14-21.

Burstein, L. (1980). Issues in the aggregation of data. In D. Berliner (Ed.), Review of Research in Education (Vol. 8). Washington, D.C.: American Educational Research Association.

Carkhuff, R. (1969). Helping and human relations: A primer for lay and professional helpers. (Volume II. Practice and research). New York: Holt, Rinehart, & Winston.

Carkhuff, R. (1971). The development of human resources. New York: Holt, Rinehart, & Winston.

Centra, J., & Potter, D. (1980). School and teacher effects: An interrelational model. Review of Educational Research, 50, 273-291.

Chang, S., & Raths, J. (1971). The schools' contribution to the cumulating deficit. Journal of Educational Research, 64, 272-276.

Clark, R. (1982). Antagonism between achievement and enjoyment in ATI studies. Educational Psychologist, 17, 92-101.

Clark, C., Gage, N., Marx, R., Peterson, P., Stayrook, N., & Winne, P. (1979). A factorial experiment on teacher structuring, soliciting, and reacting. Journal of Educational Psychology, 71, 534-552.

Clasen, D. (1983, April). The effect of four different instructional strategies on the achievement of gifted seventh grade students. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Coker, H., Medley, D., & Soar, R. (1980). How valid are expert opinions about effective teaching? Phi Delta Kappan, 62, 131-134, 149.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., & York, R. (1966). Equality of educational opportunity. Washington, D.C.: U.S. Office of Health, Education, and Welfare.

Comber, L., & Keeves, J. (1973). Science education in nineteen countries. New York: Halsted Press.

Cooley, W., & Leinhardt, G. (1980). The Instructional Dimensions Study. Educational Evaluation and Policy Analysis, 2, 7-25.

Crawford, J., Gage, N., Corno, L., Stayrook, N., Mitman, A., Schunk, D., Stallings, J., Baskin, E., Harvey, P., Austin, D., Cronin, D., & Newman, R. (1978). An experiment on teacher effectiveness and parent-assisted instruction in the third grade (Vols. 1-3). Stanford, CA: Center for Educational Research at Stanford, Stanford University.

Crawford, J. (1983). A study of instructional processes in Title I classes: 1981-82. Journal of Research and Evaluation of the Oklahoma City Public Schools, 13(1).

Cruickshank, D. (1976). Synthesis of selected recent research on teacher effects. Journal of Teacher Education, 27(1), 57-60.

Cruickshank, D., Kennedy, J., Bush, A., & Myers, B. (1979). Clear teaching: What is it? British Journal of Teacher Education, 5, 27-33.

Dalton, M., & Willcocks, J. (Eds.). (1983). Moving from the primary classroom. London, England: Routledge & Kegan Paul.

Denham, C., & Lieberman, A. (Eds.). (1980). Time to learn. Washington, D.C.: National Institute of Education.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. In L. Shulman (Ed.), Review of research in education, (Vol. 5). Itasca, IL: Peacock.

Doyle, W. (1979). Classroom task and students' abilities. In P. Peterson & H. Walberg (Eds.), Research on teaching: Concepts, findings, and implications. Berkeley, CA: McCutchan.

Doyle, W. (1983). Academic work. Review of Educational Research, 53, 159-199.

Dunkin, M. (1978). Student characteristics, classroom processes, and student achievement. Journal of Educational Psychology, 70, 998-1009.

Dunkin, M., & Biddle, B. (1974). The study of teaching. New York: Holt, Rinehart, & Winston.

Dunkin, M., & Doenau, S. (1980). A replication study of unique and joint contributions to variance in student achievement. Journal of Educational Psychology, 72, 394-403.

Eaton, J., Anderson, C., & Smith, E. (1984). Students' misconceptions interfere with science learning: Case studies of fifth grade students. Elementary School Journal, 84(4), 365-379. (Also available as IRT Research Series No. 128. East Lansing, MI: Institute for Research on Teaching, Michigan State University.)

Ebmeier, H., & Good, T. (1979). The effects of instructing teachers about good teaching on the mathematics achievement of fourth grade students. American Educational Research Journal, 16, 1-16.

Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A microanalysis of teacher-student interaction. Sociology of Education, 54, 151-173.

Emmer, E., Evertson, C., & Anderson, L. (1980). Effective classroom management at the beginning of the school year. Elementary School Journal, 80, 219-231.

Emmer, E., Evertson, C., & Brophy, J. (1979). Stability of teacher effects in junior high classrooms. American Educational Research Journal, 16, 71-75.

Engelmann, S., & Carnine, D. (1982). Theory of instruction: Principles and applications. New York: Irvington.

Evertson, C. (1979). Teacher behavior, student achievement, and student attitudes: Descriptions of selected classrooms (Report No. 4063). Austin: Research and Development Center for Teacher Education, University of Texas at Austin.

Evertson, C., Anderson, C., Anderson, L., & Brophy, J. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. American Educational Research Journal, 17, 43-60.

Evertson, C., Anderson, L., & Brophy, J. (1978). Texas Junior High School Study: Final report of process-outcome relationships (Report No. 4061). Austin: Research and Development Center for Teacher Education, University of Texas.

Evertson, C., & Brophy, J. (1973). High-inference behavioral ratings as correlates of teaching effectiveness. JSAS Catalog of Selected Documents in Psychology, 3, 97. (ERIC ED 095 174)

Evertson, C., & Brophy, J. (1974). Texas Teacher Effectiveness Project: Questionnaire and interview data (Research Report No. 74-5). Austin: Research and Development Center for Teacher Education, University of Texas at Austin. (ERIC ED 099 346)

Evertson, C., & Emmer, E. (1982). Effective management at the beginning of the school year in junior high classes. Journal of Educational Psychology, 74, 485-498.

Evertson, C., Emmer, E., & Brophy, J. (1980). Predictors of effective teaching in junior high mathematics classrooms. Journal for Research in Mathematics Education, 11, 167-178.

Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (1980). Teaching behaviors academic learning time, and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), Time to learn. Washington, D.C.: National Institute of Education.

Fisher, C., Filby, N., Marliave, R., Cahen, L., Dishaw, M., Moore, J., & Berliner, D. (1978). Teaching behaviors, academic learning time and student achievement: Final report of Phase III-B, Beginning Teacher Evaluation Study. San Francisco, CA: Far West Laboratory for Educational Research and Development.

Fitz-Gibbon, C., & Clark, K. (1982). Time variables in classroom research: A study of eight urban secondary school mathematics classes. British Journal of Educational Psychology, 52, 301-316.

Flanders, N. (1965). Teacher influence, pupil attitude, and achievement (Cooperative Research Monograph No. 12). Washington, D.C.: U.S. Office of Education.

Flanders, N. (1970). Analyzing teacher behavior. Reading, MA: Addison-Wesley.

Flanders, N., & Simon, A. (1969). Teacher effectiveness. In R. Ebel (Ed.), Encyclopedia of Educational Research (4th ed.). New York: Macmillan.

Florio, S. (1978). Learning how to go to school: An ethnography of interaction in a kindergarten-first-grade classroom. Unpublished doctoral dissertation. Cambridge, MA: Harvard University.

Fortune, J. (1967). A study of the generality of presenting behaviors in teaching preschool children (Final Report for U.S. Office of Education Project No. 6-8468). Memphis, TN: Memphis State University. (ERIC ED 016 285)

Gage, N. (1965). Desirable behaviors of teachers. Urban Education, 1, 85-95.

Gage, N. (1978). The scientific basis of the art of teaching. New York: Teachers College Press, Columbia University.

Gage, N. (1979). The generality of dimensions of teaching. In P. Peterson & H. Walberg (Eds.), Research on teaching: Concepts, findings, and implications. Berkeley, CA: McCutchan.

Gage, N. (1983). When does research on teaching yield implications for practice? Elementary School Journal, 83, 492-496.

Gage, N., Belgard, M., Dell, D., Hiller, J., Rosenshine, B., & Unruh, W. (1968). Explorations of the teacher's effectiveness in explaining (Technical Report No. 4). Stanford, CA: Stanford Center for Research and Development in Teaching, Stanford University.

Gage, N., & Coladarci, T. (1980). Replication of an experiment with a research-based inservice teacher education program. Stanford, CA: Program on Teaching Effectiveness, Center for Educational Research at Stanford, Stanford University.

Gall, M., Ward, B., Berliner, D., Cahen, L., Winne, P., Elashoff, J., & Stanton, G. (1978). Effects of questioning techniques and recitation on student learning. American Educational Research Journal, 15, 175-199.

Galton, M., & Simon, B. (1980). Progress and performance in the primary classroom. Boston, MA: Routledge & Kegan Paul.

Glass, G., & Smith, M. (1978). Meta-analysis of research on the relationship of class size and achievement. Boulder, CO: Laboratory of Educational Research, University of Colorado.

Good, T. (1979). Teacher effectiveness in the elementary school: What we know about it now. Journal of Teacher Education, 30, 52-64.

Good, T., Biddle, B., & Brophy, J. (1975). Teachers make a difference. New York: Holt, Rinehart, & Winston.

Good, T., & Brophy, J. (1984). Looking in classrooms (3rd ed.). New York: Harper & Row.

Good, T., Ebmeier, H., & Beckerman, T. (1978). Teaching mathematics in high and low SES classrooms: An empirical comparison. Journal of Teacher Education, 29, 85-90.

Good, T., & Grouws, D. (1975). Process-product relationship in fourth-grade mathematics classrooms (Final report). Columbia: College of Education, University of Missouri. NIE-G-00-3-0123

Good, T., & Grouws, D. (1977). Teaching effects: A process-product study in fourth grade mathematics classrooms. Journal of Teacher Education, 28, 49-54.

Good, T., & Grouws, D. (1979b). The Missouri Mathematics Effectiveness Project: An experimental study in fourth grade classrooms. Journal of Educational Psychology, 71, 355-362.

Good, T., & Grouws, D. (1979a). Experimental study of mathematics instruction in elementary schools (Final report). Columbia, MO: Center for the Study of Social Behavior, University of Missouri. (NIE-G-79-0103)

Good, T., & Grouws, D. (1981). Experimental research in secondary mathematics (Final report). Columbia, MO: Center for the Study of Social Behavior, University of Missouri. (NIE-G-79-0103)

Good, T., Grouws, D., & Beckerman, T. (1978). Curriculum pacing: Some empirical data in mathematics. Journal of Curriculum Studies, 10, 75-81.

Good, T., Grouws, D., & Ebmeier, M. (1983). Active mathematics teaching. New York: Longman.

Good, T., & Power, C. (1976). Designing successful classroom environments for different types of students. Journal of Curriculum Studies, 8, 1-16.

Hamilton, S. (1983). The social side of schooling: Ecological studies of classrooms and schools. Elementary School Journal, 83, 313-334.

Harris, A., & Serwer, B. (1966). The CRAFT Project: Instructional time in reading research. Reading Research Quarterly, 2, 27-57.

Heath, R., & Nielson, M. (1974). The research basis for performance-based teacher education. Review of Educational Research, 44, 463-484.

Hiller, J., Fisher, G., & Kaess, W. (1969). A computer investigation of verbal characteristics of effective classroom lecturing. American Educational Research Journal, 6, 661-675.

Hughes, D. (1973). An experimental investigation of the effects of pupil responding and teacher reacting on pupil achievement. American Educational Research Journal, 10, 21-37.

Husen, T. (Ed.). (1967). International study of achievement in mathematics (Vols. 1-2). New York: Wiley and Sons.

Janicki, T., & Peterson, P. (1981). Aptitude-treatment interaction effects of variations in direct instruction. American Educational Research Journal, 18, 63-82.

Jencks, C., Smith, M., Acland, H., Bane, M., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). Inequality: A reassessment of the effect of family and schooling in America. New York: Basic Books.

Karweit, N. (1983). Time on task: A research review (Report No. 332). Baltimore, MD: Center for Social Organization for Schools, The Johns Hopkins University.

Karweit, N., & Slavin, R. (1982). Time-on-task: Issues of timing, sampling, and definition. Journal of Educational Psychology, 74, 844-851.

Kennedy, J., Cruickshank, D., Bush, A., & Myers, B. (1978). Additional investigations into the nature of teacher clarity. Journal of Educational Research, 72, 3-10.

Kounin, J. (1970). Discipline and group management in classrooms. New York: Holt, Rinehart, & Winston.

Kyriacou, C., & Newson, G. (1982). Teacher effectiveness: A consideration of research problems. Educational Review, 34, 3-12.

Land, M. (1979). Low-inference variables and teacher clarity: Effects on student concept learning. Journal of Educational Psychology, 71, 795-799.

Land, M., & Smith, L. (1979). The effect of low inference teacher clarity inhibitors on student achievement. Journal of Teacher Education, 31, 55-57.

Larrivee, B., & Algina, J. (1983, April). Identification of teaching behaviors which predict success for mainstreamed students. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

MacKay, A. (1979). Project Quest: Teaching strategies and pupil achievement (Research Report No. 79-1-3). Edmonton: Centre for Research in Teaching, Faculty of Education, University of Alberta.

Madike, F. (1980). Teacher classroom behaviors involved in microteaching and student achievement: A regression study. Journal of Educational Psychology, 72, 265-274.

Martin, J. (1979). Effects of teacher higher-order questions on student process and product variables in a single-classroom study. Journal of Educational Research, 72, 183-187.

McCaleb, J., & White, J. (1980). Critical dimensions in evaluating teacher clarity. Journal of Classroom Interaction, 15, 27-30.

McConnell, J. (1977). Relationships between selected teacher behaviors and attitudes/achievements of algebra classes. Paper presented at the annual meeting of the American Educational Research Association.

McDonald, F. (1976). Report on Phase II of the Beginning Teacher Evaluation Study. Journal of Teacher Education, 27(1), 39-42.

McDonald, F. (1977). Research on teaching: Report on Phase II of the Beginning Teacher Evaluation Study. In G. Borich & K. Fenton (Eds.), The appraisal of teaching: Concepts and process. Reading, MA: Addison-Wesley.

McDonald, F., & Elias, P. (1976b). The effects of teaching performance on pupil learning (Final report) (Vols. 1-5), Beginning Teacher Evaluation Study, Phase II, 1974-1976. Princeton, NJ: Educational Testing Service.

McDonald, F., & Elias, P. (1976a). Executive summary report: Beginning Teacher Evaluation Study, Phase II. Princeton, NJ: Educational Testing Service.

McKenzie, G., & Henry, M. (1979). Effects of testlike events on on-task behavior, test anxiety, and achievement in a classroom rule-learning task. Journal of Educational Psychology, 71, 370-374.

Medley, D. (1977). Teacher competency and teacher effectiveness: A review of process-product research. Washington, D.C.: American Association of Colleges for Teacher Education.

Medley, D. (1979). The effectiveness of teachers. In P. Peterson & H. Walberg (Eds.), Research on teaching: Concepts, findings, and implications. Berkeley, CA: McCutchan.

Medley, D., & Mitzel, H. (1963). Measuring classroom behavior by systematic observation. In N. Gage (Ed.), Handbook of research on teaching. Chicago, IL: Rand McNally.

Mitzel, H. (1960). Teacher effectiveness. In C. Harris (Ed.), Encyclopedia of Educational Research (3rd ed.). New York: Macmillan.

Morine-Dershimer, G. (1982). Pupil perceptions of teacher praise. Elementary School Journal, 82, 421-434.

Morsh, J., & Wilder, E. (1954). Identifying the effective instructor: A review of the quantitative studies, 1900-1952 (Research Bulletin No. AFPTRIC-TR-54-44). San Antonio, TX: USAF Personnel Training Research Center, Lackland Air Force Base.

Mosteller, F., & Moynihan, D. (1972). On equality of educational opportunity. New York: Random House.

National Institute of Education. (1974). Report of Panel 2 of the National Conference on Studies in Teaching: Teaching as human interaction. Washington, D.C.: U.S. Department of Health, Education, and Welfare, National Institute of Education.

Nuthall, G., & Church, J. (1973). Experimental studies of teaching behaviour. In G. Chanan (Ed.), Towards a science of teaching. London, England: National Foundation for Educational Research.

Peterson, P. (1979). Direct instruction reconsidered. In P. Peterson & H. Walberg (Eds.), Research on teaching: Concepts, findings, and implications. Berkeley, CA: McCutchan.

Peterson, P., & Walberg, H. (1979). (Eds.). Research on teaching: Concepts, findings, and implications. Berkeley, CA: McCutchan.

Popham, W. (1971). Performance tests of teaching proficiency: Rationale, development and validation. American Educational Research Journal, 8, 105-117.

Powell, M. (1980). The Beginning Teacher Evaluation Study: A brief history of a major research project. In C. Denham & A. Lieberman (Eds.), Time to learn. Washington, D.C.: National Institute of Education.

Ramp, E., & Rhine, W. (1981). Behavior analysis model. In W. Rhine (Ed.), Making schools more effective: New directions from Follow Through. New York: Academic Press.

Redfield, D., & Rousseau, E. (1981). A meta-analysis of experimental research on teacher questioning behavior. Review of Educational Research, 51, 237-245.

Romberg, T. (1983). A common curriculum for mathematics. In G. Fenstermacher & J. Goodlad (Eds.), Individual differences and the common curriculum (82nd Yearbook of the National Society for the Study of Education, Part 1). Chicago, IL: University of Chicago Press.

Rose, J., & Medway, F. (1981). Teacher locus of control, teacher behavior, and student behavior as determinants of student achievement. Journal of Educational Research, 74, 375-381.

Rosenholtz, S., & Cohen, E. (1983). Back to basics and the desegregated school. Elementary School Journal, 83, 515-527.

Rosenholtz, S., & Wilson, B. (1980). The effect of classroom structure on shared perceptions of ability. American Educational Research Journal, 17, 75-82.

Rosenshine, R. (1968). To explain: A review of research. Educational Leadership, 26, 275-280.

Rosenshine, B. (1970a). Evaluation of instruction. Review of Educational Research, 40, 279-301.

Rosenshine, B. (1970b). Experimental classroom studies of indirect teaching. Classroom Interaction Newsletter, 5(2), 7-11.

Rosenshine, B. (1971). Teaching behaviours and student achievement. London, England: National Foundation for Educational Research.

Rosenshine, B. (1976). Classroom instruction. In N. Gage (Ed.), The psychology of teaching methods (77th Yearbook, National Society for the Study of Education). Chicago, IL: University of Chicago Press.

Rosenshine, B. (1979). Content, time, and direct instruction. In P. Peterson & H. Walberg (Eds.), Research on teaching: Concepts, findings, and implications. Berkeley, CA: McCutchan.

Rosenshine, B. (1983). Teaching functions in instructional programs. Elementary School Journal, 83, 335-351.

Rosenshine, B., & Berliner, D. (1978). Academic engaged time. British Journal of Teacher Education, 4, 3-16.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. Travers (Ed.), Second handbook of research on teaching. Chicago, IL: Rand McNally.

Rosenshine, B., & Stevens, R. (In press). Classroom instruction in reading. In D. Pearson (Ed.), Handbook of research on teaching. New York: Longman.

Rowe, M. (1974). Wait-time and rewards as instructional variables, their influence on language, logic and fate control: Part I, Wait-time. Journal of Research in Science Teaching, 11, 81-94.

Ryan, F. (1973). Differentiated effects of levels of questioning on student achievement. Elementary School Journal, 41, 63-67.

Ryan, F. (1974). The effects on social studies achievement of multiple student responding to different levels of questioning. Journal of Experimental Education, 42, 71-75.

Schofield, H. (1981). Teacher effects on cognitive and affective pupil outcomes in elementary school mathematics. Journal of Educational Psychology, 73, 462-471.

Schuck, R. (1981). The impact of set induction on student achievement and retention. Journal of Educational Research, 74, 227-232.

Simon, A., & Boyer, E. (Eds.). (1967). Mirrors for behavior: An anthology of observation instruments. Philadelphia, PA: Research for Better Schools, Inc.

Simon, A., & Boyer, E. (Eds.). (1970). Mirrors for behavior: An anthology of observation instruments continued, 1970 Supplement (Vols. A-B). Philadelphia, PA: Research for Better Schools, Inc.

Slavin, R. (1980). Cooperative learning. Review of Educational Research, 50, 315-342.

Slavin, R. (1983). Team-assisted individualization: A cooperative learning solution for adaptive instruction in mathematics (Report No. 340). Baltimore, MD: Center for the Social Organization of Schools, The Johns Hopkins University.

Smith, L. (1977). Aspects of teacher discourse and student achievement in mathematics. Journal for Research in Mathematics Education, 8, 195-204.

Smith, L. (1979). Task-oriented lessons and student achievement. Journal of Educational Research, 73, 16-19.

Smith, L., & Cotten, M. (1980). Effect of lesson vagueness and discontinuity on student achievement and attitudes. Journal of Educational Psychology, 72, 670-675.

Smith, L., & Land, M. (1981). Low-inference verbal behaviors related to teacher clarity. Journal of Classroom Interaction, 17, 37-42.

Smith, L., & Sanders, K. (1981). The effects on student achievement and student perception of varying structure in social studies content. Journal of Educational Research, 74, 333-336.

Smith, M., & Glass, G. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. American Educational Research Journal, 17, 419-433.

Soar, R. S. (1966). An integrative approach to classroom learning (Report for NIMH Projects No. 5-R11 MH 01096 and 7-R11 MH 02045). Philadelphia, PA: Temple University. (ERIC ED 033 749)

Soar, R. S. (1968). Optimum teacher-pupil interaction for pupil growth. Educational Leadership, 26, 275-280.

Soar, R. S. (1973). Follow Through classroom process measurement and pupil growth 1970-71, (Final Report). Gainesville, FL: Institute for Development of Human Resources, University of Florida. (ERIC ED 106 297)

Soar, R. S. (1977). An integration of findings from four studies of teacher effectiveness. In G. Borich & K. Fenton (Eds.), The appraisal of teaching: Concepts and process. Reading, MA: Addison-Wesley.

Soar, R. S., & Soar, R. M. (1972). An empirical analysis of selected Follow Through programs: An example of a process approach to evaluation. In I. Gordon (Ed.), Early childhood education. Chicago, IL: National Society for the Study of Education.

Soar, R. S., & Soar, R. M. (1973). Classroom behavior, pupil characteristics and pupil growth for the school year and the summer. Gainesville, FL: Institute for Development of Human Resources, University of Florida.

Soar, R. S., & Soar, R. M. (1978). Setting variables, classroom interaction, and multiple pupil outcomes (Final Report, Project No. 6-0432). Washington, D.C.: National Institute of Education Grant No. (NIE-G-76-0100)

Soar, R. S., & Soar, R. M. (1979). Emotional climate and management. In P. Peterson & H. Walberg (Eds.), Research on teaching: Concepts, findings, and implications. Berkeley, CA: McCutchan.

Solomon, D., & Kendall, A. (1979). Children in classrooms: An investigation of person-environment interaction. New York: Praeger.

Southwest Educational Development Laboratory. (1973). Bilingual kindergarten program inservice manual (Vol. 1). Austin, TX: National Educational Laboratory.

Stallings, J. (1975). Implementation and child effects of teaching practices in Follow Through classrooms. Monographs of the Society for Research in Child Development, 40, (7-8 Serial No. 163).

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. Educational Researcher, 8,(11), 11-16.

Stallings, J., Cory, R., Fairweather, J., & Needels, M. (1977). Early childhood education classroom evaluation. Menlo Park, CA: SRI International.

Stallings, J., Cory, R., Fairweather, J., & Needels, M. (1978). A study of basic reading skills taught in secondary schools. Menlo Park, CA: SRI International.

Stallings, J., & Kaskowitz, D. (1974). Follow Through classroom observation evaluation, 1972-1973. Stanford, CA: Stanford Research Institute. (SRI Project URU-7370).

Stallings, J., Needels, M., & Stayrook, N. (1979). The teaching of basic reading skills in secondary schools, Phase II and Phase III. Menlo Park, CA: SRI International.

Tikunoff, W., Berliner, D., & Rist, R. (1975). An ethnographic study of the forty classrooms of the Beginning Teacher Evaluation Study known sample (Technical Report No. 75-10-5). San Francisco, CA: Far West Laboratory for Educational Research and Development.

Tobin, K. (1980). The effect of an extended teacher wait-time on science achievement. Journal of Research in Science Teaching, 17, 469-475.

Tobin, K., & Capie, W. (1982). Relationships between classroom process variables and middle-school science achievement. Journal of Educational Psychology, 74, 441-454.

Veldman, D., & Brophy, J. (1974). Measuring teacher effects on pupil achievement. Journal of Educational Psychology, 66, 319-324.

Walker, D., & Schaffarzick, J. (1974). Comparing curricula. Review of Educational Research, 44, 83-111.

Webb, N. (1980). A process-outcome analysis of learning in group and individual settings. Educational Psychologist, 15, 69-83.

Weinstein, R. (1983). Student perceptions of schooling. Elementary School Journal, 83, 287-312.

Winne, P. (1979). Experiments relating teachers' use of higher cognitive questions to student achievement. Review of Educational Research, 49, 13-50.

Winne, P., & Marx, R. (1982). Students' and teachers' views of thinking processes involved in classroom learning. Elementary School Journal, 82, 493-518.

Wright, C., & Nuthall, G. (1970). Relationships between teacher behaviors and pupil achievement in three experimental elementary science lessons. American Educational Research Journal, 7, 477-491.

Wyne, M., & Stuck, G. (1982). Time and learning: Implications for the classroom teacher. Elementary School Journal, 83, 67-75.

APPENDIX

Revised Principles for Small-Group Instruction in Beginning Reading

Anderson, Evertson, and Brophy (1982)

General Principles

1. Reading groups should be organized for efficient, sustained focus on the content.

2. All students should be not merely attentive but actively involved in the lesson.

3. The difficulty level of questions and tasks should be easy enough to allow the lesson to move along at a brisk pace and the students to experience consistent success.

4. Students should receive frequent opportunities to read and respond to questions and should get clear feedback about the correctness of their performance.

5. Skills should be mastered to overlearning, with new ones gradually phased in while old ones are being mastered.

6. Although instruction takes place in the group setting, monitor each individual and provide whatever instruction, feedback, or opportunities to practice that he or she requires.

Specific Principles

Programming for Continuous Progress

1. Time. Across the year, reading groups should average 25-30 minutes each. The length will depend on student attention level, which varies with time of year, student ability level, and the skills being taught.

2. Academic focus. Successful reading instruction includes not only organization and management of the reading group itself (discussed below), but effective management of the students who are working independently. Provide these students with (1) appropriate assignments, (2) rules and routines to follow when they need help or information (to minimize their needs to interrupt you as you work with your reading group), (3) and activity options for when they finish their work (so they have something else to do).

3. Pace. Both progress through the curriculum and pacing within specific activities should be brisk, producing continuous progress achieved with relative ease (small steps, high success rate).

4. Error rate. Expect to get correct answers to about 80% of your questions in reading groups. More errors can be expected when students are working on new skills (perhaps 20-30%). Continue with practice and review until smooth, rapid, correct performance is achieved. Review responses should be almost completely (perhaps 95%) correct.

Organizing the Group

5. Seating. Arrange seating so that you can both work with the reading group and monitor the rest of the class at the same time.

6. Transitions. Teach the students to respond immediately to a signal to move into the reading group (bringing their books or other materials), and to make quick, orderly transitions between activities.

7. Getting started. Start lessons quickly once the students are in the group (have your materials prepared beforehand).

Introducing Lessons and Activities

8. Overviews. Begin with an overview to provide students with a mental set and help them anticipate what they will be learning.

9. New words. When presenting new words, do not merely say the word and move on. Usually, you should show the word and offer phonetic clues to help students learn to decode.

10. Work assignments. Be sure that students know what to do and how to do it. Before releasing them to work on activities independently, have them demonstrate how they will accomplish these activities.

Insuring Everyone's Participation

11. Ask questions. In addition to having the students read, ask them questions about the words and materials. This helps keep students attentive during classmates' reading turns and allows you to call their attention to key concepts or meanings.

12. Ordered turns. Use a system, such as going in order around the group, to select students for reading or answering questions. This insures that all students have opportunities to participate, and it simplifies group management by eliminating handwaving and other student attempts to get you to call on them.

13. Minimize call outs. In general, minimize student call-outs and emphasize that students must wait their turns and respect the turns of others. Occasionally, you may want to allow call-outs to pick up the pace or encourage interest, especially with low achievers or students who do not normally

volunteer. If so, give clear instructions or devise a signal to indicate that you intend to allow call-outs at these times.

14. Monitor individuals. Be sure that everyone, but especially slow students, is checked, receives feedback, and achieves mastery. Ordinarily this will require questioning each individual student, and not relying on choral responses.

Teacher Questions and Student Answers

15. Academic focus. Concentrate your questions on the academic content; do not overdo questions about personal experiences. Most questions should be about word recognition or sentence or story comprehension.

16. Word attack questions. Include word attack questions that require students to decode words or identify sounds within words.

17. Wait for answers. In general, wait for an answer if the student is still thinking about the question and may be able to respond. However, do not continue waiting if the student seems lost or is becoming embarrassed or if you are losing the other students' attention.

18. Give needed help. If you think the student cannot respond without help but may be able to reason out the correct answer with help, provide help by simplifying the question, rephrasing the question, or giving clues.

19. Give the answer when necessary. When the student is unable to respond, give the answer or call on someone else. In general, focus the attention of the group on the answer, not on the failure to respond.

20. Explain the answer when necessary. If the question requires one to develop a response by applying a chain of reasoning or step-by-step problem solving, explain the steps one goes through to arrive at the answer in addition to giving the answer itself.

When the Student Responds Correctly

21. Acknowledge correctness (unless it is obvious). Briefly acknowledge the correctness of responses (nod, repeat the answer, say "right," etc.), unless it is obvious to the students that their answers are correct (such as during fast-paced drills reviewing old material).

22. Explain the answer when necessary. Even after correct answers, feedback that emphasizes the methods used to get answers will often be appropriate. Onlookers may need this information to understand why the answer is correct.

23. Follow up questions. Occasionally, you may want to address one or more follow up questions to the same student. Such series of related questions can help the student to integrate relevant information. Or, you may want to extend a line of questioning to its logical conclusion.

Praise and Criticism

24. Praise in moderation. Praise only occasionally (no more than perhaps 10% of correct responses). Frequent praise, especially if nonspecific, is probably less useful than more informative feedback.

25. Specify what is praised. When you do praise, specify what is being praised, if this is not obvious to the student and the onlookers.

26. Correction, not criticism. Routinely inform students whenever they respond incorrectly, but in ways that focus on the academic content and include corrective feedback. When it is necessary to criticize (typically only about 1% of the time when students fail to respond correctly), be specific about what is being criticized and about desired alternative behaviors.