

# Recent Research on Teaching Behaviors and Student Achievement <sup>1</sup>

Barak Rosenshine

(University of Illinois at Urbana-Champaign)

Although studies on teacher behaviors and teacher effects have been reported since 1940, the modern era of this research began in 1957 with the work of investigators such as Flanders, Medley, and Mitzel. Even since then, the number of studies has been small. Fewer than 25 studies have been conducted on any specific variable such as teacher praise or teacher questions, and these studies are spread across all grade levels, subject areas, and student backgrounds.

The number of investigators in this field is also small. There are no more than 12 researchers <sup>2</sup> or groups of researchers currently studying the relationship between classroom instruction and student achievement. These 12 are spread out across the different instructional contexts and variables being studied. The number of researchers actively collecting data in any given year is much smaller than 12. Furthermore, these researchers have met only twice, as a group, to discuss common findings, inconsistencies, and problems.

Although the recent studies, summarized in this issue, represent methodological and conceptual expansion of previous work, research on observed teaching behavior is new, sparse, and not always consistent in results. What we have learned to date is offered more as hypotheses for future study than as validated variables for the training and evaluation of teachers. **Although practitioners can easily amass a large number of questions on teaching methods for which they would like clear answers, at the rate we are going it will be years before many of these questions are even studied.**

---

<sup>1</sup> . *Journal of Teacher Education*, Spring 1976, vol. 27 n° 1, pp. 61-64.

<sup>2</sup> . The 12 or so researchers, groups of researchers, and their organizations are:

- David Berliner, Charlie Fisher, and Len Cahen – Far West Laboratory;
- Walter Borg – Utah State University;
- Jere Brophy and Carolyn Evertson – Texas Research and Development Center in Teacher Education;
- Homer Coker – Carrolton State College, Georgia, and Don Medley – University of Virginia;
- N.L. Gage – Stanford Research and Development Center in Teaching;
- Meredith Gall, William Tikunoff, and Betty Ward – Far West Laboratory;
- Tom Good – University of Missouri;
- Gene Hall, Sue Loucks, Gary Borich, and Robert Peck – Texas Research and Development Center in Teaching;
- Gaea Leinhardt, Margaret Wang, and William Cooley – Learning Research and Development Center, University of Pittsburg;
- Fred McDonald – Educational Testing Service;
- Robert Soar – University of Florida;
- Jane Stallings – Stanford Research Institute.

In 1975, a number of these researchers met in San Diego and again in Austin to present and discuss their recent research, and a number of these papers are summarized below. Unfortunately, these papers do not fit into one piece. Each investigator worked on a problem individually, and there has been little communication between them. Hopefully, now that people have been brought together to share ideas, there can be more coordination and cooperation.

## Summary of Recent Studies

The topics of these papers are too different to permit easy grouping. Yet, I intuitively grouped them into categories: experimental studies of questioning, tutoring, program implementation, generic and specific teaching skills, and teaching low SES students.

### *Experimental Studies of Questioning*

Teacher questions have been a major area in all teacher training programs, yet the research base for our practice is woefully thin. The two experimental studies of questioning – by Ward and Tickunoff, 1975<sup>3</sup> and Stanford Program on Teaching Effectiveness<sup>4</sup> – are exciting because they represent serious and well-designed attempts to overcome design deficiencies of previous studies.

Overall, both studies found that classes where students were asked more recall questions did slightly better on the recall tests, whereas all classes did equally well on integrative questions, no matter what percent were asked in the class. In an additional analysis reported in Ward and Tickunoff, it was found that low ability students did best with factual questions, and without probing and redirection, whereas high ability students did best with probing and redirection.

Thus, conclusions emerge from these studies. First, factual questions appear to be functional for low ability students, a finding which is also supported by recent correlational studies<sup>5</sup>, and are not necessarily dysfunctional for high ability students. Second, we are as yet unaware of the optimal types and sequencing of questions and responses for teaching the ability to draw inferences or apply learning. Finally, the optimal sequences may be different for different types of students and outcomes, but we still do not know what these sequences are. Put another way, **the continual bromides that factual questions are bad and higher level questions are good, were not supported by well-designed research.**

### *Experimental Studies of Tutoring*

The Far West Laboratory studies of tutoring<sup>6</sup> were attempts to validate their mini-course on mathematics tutoring. The results suggested that training in tutoring, or tutoring itself, may not be as important as expected. If students are given an extra half hour a day of mathematics instruction, they learn more, regardless of whether this instruction takes place in regular class or in tutoring sessions, and regardless of whether the teachers are trained

<sup>3</sup> . B.A. Ward and W.J. Tikunoff, *Application of Research to Teaching, Report A75-2* (San Francisco, Calif.: Far West Laboratory for Educational Research, 1975).

<sup>4</sup> . Stanford Program on Teaching Effectiveness, *A Factorially Designed Experiment on Teacher Structuring, Soliciting, and Reacting* (Stanford, Calif.: Stanford Center for Research and Development in Teaching, 1976).

<sup>5</sup> . J.A. Stallings and D.H. Kaskowitz, "A Study of Follow Through Implementation", Paper presented to the annual meeting of the American Educational Research Association, April 1975 (Menlo Park, Stanford Research Institute).

J.E. Brophy and C.M. Evertson, *Process-Product Correlations in the Texas Teacher Effectiveness Study: Final Report* (Austin, Tex.: The University of Texas, 1974).

R.S. Soar, *Follow Through Classroom Process Measurement and Pupil Growth (1970-71): Final Report* (Gainesville, Fla.: College of Education, University of Florida, 1973).

<sup>6</sup> . Ward and Tikunoff, op. cit.

B.J. Williams et al., *Math Tutoring Study* (San Francisco, Calif.: Far West Laboratory for Educational Research and Development, 1975).

or untrained in tutoring. A similar study, but one which only involved 30 minutes of teaching or tutoring found that tutoring yielded significantly greater achievement than classroom teaching <sup>7</sup>. Although this study raises questions about common practice, it remains for future research to tell us effective procedures for improving tutoring or teaching.

### ***Levels of Implementation***

As research is extended and findings become incorporated into curriculum packages, the monitoring of curriculum implementation becomes important. Comparisons of curriculum packages and regular programs are not particularly useful if one does not know how well the curriculum is being implemented. Both Stallings and Kaskowitz and Soar used observational data to measure curriculum implementation. Hall and his associates <sup>8</sup> complement this work with their development of a teacher checklist (Stages of Concern) and an interview schedule (Levels of Use) to obtain information on the thought and practice of teachers who are implementing an innovation. The Levels of Use Instrument <sup>9</sup> demonstrated that, as teachers gain experience, they shift from mechanical use of an innovation to stabilized and refined use. This instrument might be useful for monitoring implementation.

### ***Generic and Specific Skills***

Two studies – the ethnographic study of the Far West Laboratory (FWL) <sup>10</sup> and the observational study by Educational Testing Service (ETS) <sup>11</sup> – are examples in which researchers developed their variables after they had recorded the events which occurred in observed classrooms. A second similarity was that both studied teaching in four contexts: reading and mathematics in both the second and fifth grades. Finally, in both studies, no differentiation was made for student SES background.

One question which the FWL study addressed is whether teaching skills are generic or specific. Overall, the FWL study found that over a third of the variables, or dimensions, which they studied were generic, that is, these 21 dimensions were significant in each of the four contexts. In addition, these researchers found that reading and mathematics were more similar at second grade than at fifth, and that across the grades reading was more similar than math. There were no inconsistent dimensions, that is, dimensions which were significantly positive in one context but negative in another.

In sharp contrast, the ETS study <sup>12</sup> found no dimensions which were generic, found that reading and math were most similar at fifth grade, and found that across the grades math was more similar than reading. Finally, there were almost as many inconsistent dimensions as consistent dimensions.

Thus, these two studies give different pictures of generic and specific skills. The FWL study argues for a mixture of generic and specific skills, with skills in second-grade reading most general and skills in fifth-grade mathematics most specific. In contrast, the ETS study presents the case that effective skills are highly unique in each context.

---

<sup>7</sup> . R.B. Bausell, W.B. Moody, and F.N. Walzl, "A Factorial Study of Tutoring Versus Classroom Instruction", *American Educational Research Journal* 9(1972): 591-99.

<sup>8</sup> . G.E. Hall, *The Effects of "Change" on Teachers and Professors — Theory, Research, and Implications for Decision Makers* (Austin, Tex.: Research and Development Center for Teacher Education, the University of Texas, 1975).

<sup>9</sup> . Gene E. Hall et al., "Levels of Use of the Innovation: A Framework for Analyzing Innovation Adoption", *Journal of Teacher Education* XXVI, no. 1 (Spring 1975): 52-56.

<sup>10</sup> . W.J. Tikunoff, D. Berliner, and R. Rist, *An Ethnographic Study of Forty Classrooms*, Technical Report no. 75-10-5 (San Francisco, Calif.: Far West Laboratory for Educational Research, 1975).

<sup>11</sup> . F.J. McDonald, et al., *Beginning Teacher Education Study, Phase II: Final Report* (Princeton, N.J.: Educational Testing Service, 1975).

<sup>12</sup> . The results of generic and specific variables in the ETS study came from my reading of a draft of their final report. McDonald (1975) did not discuss generic and specific results directly, but the lack of generic results is clear in that paper.

Such differing results are puzzling. One can argue that differences in coding procedures, selection of sample, and length of instruction are so large that comparing these studies is meaningless. At any rate, **additional studies are essential before we are clear about which skills are generic, grade level specific, or subject area specific.**

### ***Studies of Primary Grade Low SES Students***

Three of the investigations<sup>13</sup> covered the same context – primary grade reading and mathematics for children from low SES backgrounds. All three conducted correlational studies between prespecified instructional variables and student achievement. There was an exciting convergence of significant results across these three studies. Consistent positive correlations were obtained for direct time, narrow questions, teacher positive feedback, student attention to task, and supervised student study in small or large groups. The results on eight common variables in these studies are summarized in Table 1.

#### **TIME**

In these studies, the amount of time spent directly on instruction was significantly related to student achievement. For example, Stallings and Kaskowitz coded the *observed time spent* on reading and mathematics activities, and obtained significant and positive results. In addition, all three investigators coded time spent on nonacademic activities (such as dramatic play, games, or questions about home and family) and all three obtained consistent negative correlations.

Other recent studies also support the importance of direct time; Wiley and Harnischfeger<sup>14</sup> found that the average number of hours of schooling per year was positively and significantly related to achievement in reading and mathematics directly or indirectly. In other studies, the coding of *content covered* has yielded significant positive correlations with achievement<sup>15</sup>. Similarly, in a review of studies comparing different curriculum programs, Walker and Schaffarzick<sup>16</sup> concluded that the outcomes of different curriculum programs mirrored the content that was taught. That is, new curriculums were more effective when the post-test measured content relevant to them, whereas traditional curriculums were slightly superior when traditional post-tests were used.

Overall, then, both direct instructional time and content covered have been positively related to achievement, whereas time on nonacademic activities has been negatively related.

#### **QUESTIONS**

In all three studies, the frequency of factual, single-answer questions was correlated positively and significantly with achievement, whereas the frequency of more complex, difficult, or divergent questions had negative correlations. Brophy and Evertson also found that for low SES students, the percentage of correct answers was positively and significantly related to achievement. The results on questions suggest that for this context (primary grade students from low SES backgrounds) it may be preferable to proceed in small steps and ask factual questions at the child's level.

The above results, on the functional value of factual questions and undemonstrated value for higher order questions, parallel those obtained in the two cited experimental studies. **The lack of significant results for complex or higher level questions has puzzled all the researchers**, and led us to conclude that **we need to rethink what is meant by types of questions and their effects.**

<sup>13</sup> . Brophy and Evertson, op. cit. ; Soar, op. cit. ; Stallings and Kaskowitz, op. cit.

<sup>14</sup> . D.E. Wiley and A. Harnischfeger, "Explosion of a Myth: Quantity of Schooling and Exposure to Instruction, Major Educational Vehicles", *Educational Researcher* 3 (1974): 7-12.

<sup>15</sup> . Rosenshine and Furst, op .cit.

<sup>16</sup> . D. Walker and J. Schaffarzick, "Comparing Curricula", *Review of Educational Research* 44(1974): 83-111.

**TABLE 1**  
**SUMMARY OF RESULTS ON COMMON VARIABLES IN RECENT STUDIES ON PRIMARY**  
**GRADE LOW SES IN READING & MATH**

Variable	Stalling Kaskowitz	Soar	Brophy Evertson	Others
Direct time on academic activities	+			+
Time on noncurricular activities	—	—	—	
Allotted time for school or instruction			0	+
Content covered				+
Direct, narrow questions	+	+	+	+
Higher order, open questions	—	—	—	0
Student attention to task	+	0, +	0	0, +
Student inattention, misbehavior		—	—	—
Student in large group	+	+		
Students in independent study without teacher	—	—		
Students in supervised independent study		+		
Praise, & adult positive feedback	+	+	0, +	0
Criticism, & adult negative feedback	+	—	0, —	—
Accepting student comments	+		+	
Student comments-relevant	0	0	+	
Student comments-irrelevant	—	0	0	
Student questions-relevant	—	0	0	
Student questions-irrelevant	—	0	—	
Teacher initiation		—, +	—, 0	
Teacher management requests, requests for order		0	—	

+ = positive & significant correlations  
0 = nonsignificant and mixed correlations  
— = negative and significant correlations

#### STUDENT INATTENTION

Student inattention was consistently, significantly, and negatively related to achievement in all three studies. Similar negative results have usually also been obtained by McDonald<sup>17</sup> and other investigators. The results for student attention or on-task behavior were positive, but the correlations were not as high or consistent as those for inattention; again, similar results were obtained in other studies.

<sup>17</sup> . McDonald, et al., op. cit.

### WORK GROUPINGS

For both Stallings and Kaskowitz, and for Soar, positive and significant correlations were obtained for students *working in groups* or doing seatwork under supervision. Both investigators also found *negative* correlations for children working independently without supervision. The researchers agreed that **independent study without supervision does not appear successful unless students are first taught how to work independently.**

### ADULT FEEDBACK, PRAISE AND CRITICISM

Overall, teacher praise showed consistent, positive, but low correlations with student achievement. Praise of student academic responses had higher correlations than praise for student behavior. However, the results were not consistent for academic criticism — criticism following a student answer. Stallings and Kaskowitz found that criticism following an incorrect answer has positive correlations with achievement, whereas Brophy and Evertson found significant negative correlations for the same variable, and Soar obtained negative-but-low correlations.

### STUDENT-INITIATED AND TEACHER-INITIATED COMMENTS

With one exception all types of student-initiated talk, whether academic or nonacademic, yielded negative or low correlations. In discussing this, researchers concluded that student-initiated talk does not appear to be as important for this type of achievement as once thought.

The frequency of teacher-initiated comments did not show consistent results in these studies – sometimes yielding significant positive and sometimes significant negative correlations. In discussing this finding, the researchers suggested that coding the probable cause of a teacher's comment (e.g., management, clarification, feedback, or motivation) might be helpful in future research.

### MANAGEMENT AND CONTROL REQUESTS

There were no consistent results for teacher management or control statement. Brophy and Evertson found that such statements (e.g., “silence” “get to work”) were negatively related to achievement. However, Stallings and Kaskowitz found significant *positive* correlations (in third grade) both for adult punishment of children and for negative corrective feedback on academic matters, while Soar found nonsignificant relationships between control statements and achievement.

## ***Direct Instruction for Low SES Pupils***

In the three studies on teaching low SES students in primary grades, there was convergence on an optimal pattern for this instruction, which might be labeled as *direct instruction*. In direct instruction a great deal of time is spent on academic activities, with a predominance of seatwork using structured materials. Teacher and workbook questions are narrow and direct, usually with a single correct answer. Teachers or materials provide immediate feedback using praise and acknowledgement of student answers. Students work in groups supervised by the teacher with little free time or unsupervised activity, resulting in less off-task student behavior.

The results of some studies also suggest that in direct instruction, the teacher is the dominant leader who decides which activities will take place. The learning is approached in a direct business-like manner and is organized around questions posed by the teacher or the materials. Materials and instruction are systematically organized, proceeding within small steps. Goals are clear and known to the students. Yet, within this task setting the teacher is warm and convivial, frequently giving praise and encouragement to the students for academic work. Future investigations will be needed to generalize these findings, both for low SES and middle SES students.

## Overall

The recent studies covered a number of topics: questioning, tutoring, implementing innovations, generic and specific skills, and teaching low SES students. The studies on questioning and tutoring serve mainly to raise questions about current practice while leaving answers to future research. The studies on generic teaching skills provide a fascinating but unresolved puzzle. The studies of teaching low SES students yield a convergence on a pattern of variables which was labeled direct instruction and such a pattern awaits replication and modification. One is cautiously optimistic about some of the results; yet, the rate at which this research is conducted is very slow and answers may not be immediately forthcoming.

## BIBLIOGRAPHY

- Flanders, N.A.** *Analyzing Classroom Behavior*, New York: Addison-Wesley, 1970.
- Gall, M. et al.** *The Effects of Teacher Use of Questioning Techniques on Student Achievement and Attitudes*. San Francisco, Calif.: Far West Laboratory for Educational Research and Development, 1975.
- Good, T.L. and Grouws, D.A.** *Process Product Relationships in 4th Grade Mathematics Classes*. Columbia, Mo.: College of Education, University of Missouri, 1975.
- Rosenshine, B. and Martin, M.** "Teacher Education and Teacher Behavior". *Educational Researcher* 3 (1974): 11-14.
- Stallings, J.A. and Kaskowitz, D.H.** *Follow Through Classroom Observation Evaluation (1972-73)*. Menlo Park, Calif.: Stanford Research Institute, 1974.
- Tallmadge, K.** *The Development of Project Information Packages for Effective Approaches in Compensatory Education*. Technical Report UR-254. Mountain View, Calif: RMC Research Corp., 1974.