

A Study of the Generalizability of Teacher Change Quasi-Experiments

Samuel Stringfield

Northwest Regional Education Laboratory

Eugene Schaffer

University of North Carolina at Charlotte

Roberta Devlin-Scherer

National College of Education

ABSTRACT

Quasi-experimental teacher effectiveness studies have indicated that properly designed staff development programs can lead to changes in teacher and student behavior and to gains in student achievement. Those studies involved workshop series led by nationally known scholars. Using instructors from varied backgrounds, the current study examined the extent to which the findings of prior research would generalize to contexts where experimenter effects were less likely to confound findings. Four series of the Stallings Effective Use of Time Program were analyzed for levels of teachers' classroom behavioral change. The leaders of each series were, respectively, an external consultant, university professors, local principals, and local teachers. Classroom observational data gathered before and after the four workshop series provided significant, though limited, support for the generalizability of change from such programs. The leader characteristics variable produced no significant effects, thus increasing the support for potential cost-effective use of prior teacher effectiveness quasi-experiments.

INTRODUCTION

Building on prior observational and correlational studies, several recent quasi-experiments (e.g., Crawford et al., 1978; Anderson, Evertson, & Brophy, 1979; Anderson, Evertson & Brophy, 1979; Good & Grouws, 1979; Stallings, Needels, & Stayrook, 1979; Coladarci & Gage, 1984) investigated the extent to which teacher behavior and student achievement can be altered. Four of the five studies produced statistically significant changes in teacher and student behavior, as well as student achievement gain. Only the deliberately minimal intervention Coladarci and Gage (1984) study produced no significant change.

In other quasi-experiments, change in teacher and/or student behavior has been the criterion for effectiveness. These have been similarly successful in producing change in teacher behavior through intervention.

While agreeing with Good and Brophy's (1986) cautiously optimistic review of the teacher effectiveness literature, Gage (1985) noted a limitation of the above studies:

...all of these...training efforts were conducted in the context of research enterprises, or experiments. It is not known whether similar training would be effective in the context of regular, routine staff development in schools. (pp. 34-35)

An analysis of a local implementation of the Stallings Effective Use of Time Program (EUTP) in Washington DC (Bush, 1985; Anderson, 1985), while generally replicating and extending the Stallings findings, found that the external consultant proved more effective at inducing change in teacher behavior than did in-house team leader trainees. The Anderson analysis indicated that the additional level of effectiveness appeared to be independent of actual differences between the behavior patterns of the consultant and local teacher trainers. Implicit in the finding is a strong experimenter effect (Rosenthal, 1976) seriously confounding program effect.

If replicated, that data would call into question the magnitude of program-induced change in previous teacher effectiveness quasi-experiments, and would place substantial limitations on the practical interpretation of those studies.

The current research was conducted to examine the generalizability of effect of the Stallings program. Two aspects of EUTP's generalizability were tested simultaneously. Data were gathered on EUTP workshop series conducted by four distinctly differing trainer groups. The trainer groups were chosen to be representative of the diverse population of persons commonly leading inservice training programs in education. The first series was led by an external consultant who was also involved in the Washington D.C. replication.

The second was led by two local university professors, the third by two local principals, and the fourth by a local teacher and a half-time teacher/half-time gifted education coordinator. Nationally, external consultants, university faculty members, local administrators, and teachers provide the great bulk of inservice training to teachers. These four configurations of trainers, conducting overlapping workshop series, provided the opportunity to study the generalizability across trainer conditions of the Stallings program.

Second, the data were gathered in workshops conducted in two small town-to-town rural communities in the southern United States. No prior documentation of the effectiveness of the EUTP in relatively rural contexts has been published.

STALLINGS EFFECTIVE USE OF TIME PROGRAM

Detailed descriptions of the practical workings of the Stallings program have been published elsewhere (e.g., Stallings, Needels, & Stayrook, 1978; Stallings, 1980; Devlin-Scherer, 1984; Bush, 1985; Anderson, 1985). The description of the EUTP here will be brief.

The Stallings program is a Joint Dissemination and Review Panel approved National Dissemination Network (NDN) workshop series. The goal of EUTP is to increase student achievement by changing teachers' in-class behavior patterns to more nearly resemble those of highly effective teachers. An operational definition of effective behavior patterns was derived through analysis of prior quantitative teacher effectiveness studies (e.g., Stallings & Kaskowitz, 1974).

There are three essential features to the Stallings program. The first is direct, quantified classroom observation. Second is the workshops themselves, and third is the training and certification of local EUTP leaders to continue and expand the program after the specialized consultants have completed their assignments.

THE STALLINGS OBSERVATION SYSTEM

Each teacher involved in the EUTP is observed for three one hour periods before and again after the workshop series. Data is gathered using the Stallings Observation System (SOS).

The SOS is a complex, low inference observation system designed to be sensitive to differing instructional methods, interpersonal interactions, and classroom environments. Using the SOS, an observer alternately gathers data on the teacher and his/her immediate environment and on the entire class.

In its present form or with minor modifications, the SOS has been a major process data gathering instrument in several teacher and school effectiveness studies (e.g., Stallings & Kaskowitz, 1974; Stallings, 1980; Goodlad, 1983; Stringfield, Teddlie, & Suarez, 1985).

Data from two sections of the SOS were analyzed in this study. The first section, labeled the Classroom Snapshot (CS), yields data on the activities of each adult and student in a classroom at a given moment. Further, it contains codings for size of groups and types of materials being used. The snapshot is designed to capture data on 13 activities (e.g., reading out loud), by eight types of materials (e.g., workbook), by four types of actors (e.g., teachers, individual students), by four divisions of recipients (e.g., small group).

The teacher activity analysis of CS yielded 12 variables. Used for clinical training purposes with individual teachers, these categories are discussed separately with each teacher. However, for research purposes, the fact that the frequency of occurrence of some of the categories is very low produces instability in some analyses. (For an examination of inter-rater agreement on individual categories of the SOS in slightly modified form, see Sirotnik [1983]. For four additional reliability studies, see Stallings & Kaskowitz, [1974].) In the present paper, notes will be made of individual items, but statistical analyses of CS data will be largely restricted to two aggregated CS variables: Interactive Teaching (INT) and Organizing/Off-Task activities (ORGOF).

CS variables which were aggregated to produce INT included reading aloud, instruction and explanation of new materials, review and discussion, and practice drill. Stallings (1980) found Interactive Teaching to be both a positive predictor of student achievement gain and an alterable variable.

Passing out papers and lining up students for recess are examples of organizing activities. Off-task activities include social interactions and negative interactions (e.g., discipline). Spending a large percentage of class time in ORGOF activities has been found to be negatively correlated with achievement.

A second SOS subsystem is the Five Minute Interaction (FMI). The FMI consists of a series of frames in which each teacher behavior and/ or interaction is coded into four categories: who (e.g., teacher), to whom (e.g., small group), what (e.g., asks a direct question), and how (e.g., with positive emotion). During a FMI, a minimum of one frame is recorded every six seconds, yielding 50 or more frames per five minute observation period. The analysis produces 19 separate FMI variables.

The FMI variable studied in this paper is All Academic Statements (ALACS). It is an aggregated variable and includes all academically related talking by teachers to students (individually or in groups) and students' academic talking to teachers. ALACS has been found by Stallings (1980) to be both alterable and positively related to achievement.

EUTP WORKSHOPS

The second feature of the Stallings program is a series of five highly interactive, small group workshops. To facilitate interaction among participants, group size is limited to between six and eight teachers plus one or two persons functioning as leaders/teacher trainers. A separate topic is covered at each workshop meeting.

At the first meeting, teachers are given quantified "profiles" of their behavior patterns as measured on the SOS during the classroom visits. Each individual profile contains three types of information on 33 discrete variables. The variables are specific teacher behaviors which prior research has shown to be related to student achievement gain. For each variable, each teacher is provided with normative data, a criterion performance level for effective teaching, and their own frequency of behavior on that dimension. A typical profile would indicate that a teacher was performing at or above criterion on several dimensions, yet below criterion on others.

The first EUTP meeting is spent providing an overview of the teacher effectiveness research findings, explaining the SOS profiles and their relationship to the research literature, distributing reading materials, and inviting the teachers to choose one or two variables on which they would like to experiment with changing during the following week.

The second meeting, typically held one week later, is spent reviewing the SOS profiles, discussing the prior week's efforts at change, and examining both the teaching of reading in the content areas and grouping strategies. The third workshop focuses on suggestions for organizing and structuring classroom activities. The effective management of class time is a major topic.

During the fourth weekly meeting, teachers examine studies of the management of student behavior. The trainer's assignment is to lead an (often lively) discussion of theoretical and practical knowledge of the topic within the local context. Techniques for motivating students to strive for desirable behavior and achievement are stressed. At the fifth workshop, direct instruction, the monitoring of student activity during classes, and feedback to students are discussed.

Following the fifth meeting, three additional hours of SOS data are gathered. A final meeting is held in which teachers receive a printout comparing their pre- and post- workshop behavior profiles. On the printout, areas of improvement are noted, as are areas in which continued effort might prove fruitful.

TRAINING LOCAL EUTP LEADERS

Of the previously mentioned teacher change studies, only the deliberately minimal intervention study provided teachers with no workshops. Coladarci and Gage (1984) concluded that, "[Personal] contact might well be a necessary condition for successful implementation and, consequently, associated change in targeted outcomes" (p. 551).

For such personal contact to be cost effective, it must be available on a local basis. The Stallings program has accommodated this reality by building the training of local leaders into the program. The training involves a three step process. First, the local leader trainees read the EUTP materials and observe workshops conducted by a leader-trainer previously certified by Stallings. Second, they are trained the use and interpretation of the SOS. This step typically requires seven full days. Third, those trainees conduct workshops

series themselves. Those sessions are audiotape recorded and each week the certified leader/trainer listens to and discusses sections of the tapes with the trainee. Following this method, it is possible over time to geometrically expand the number EUTP leaders in a school system.

METHOD

During the winter and spring of 1982, four series of EUTP workshop series were conducted in two small, adjoining North Carolina school systems. The systems served towns with populations under 40,000 plus residents of outlying rural areas. Each district included four elementary, one middle, and one senior high schools.

Both teachers and trainers served as subjects in the study. The 27 teachers who volunteered to participate in the NDN, sponsored program were drawn from all grades and taught subjects as diverse as general first grade and high school social studies. A single section of the Stallings program had been offered the year before in one of the school systems, and many of the teachers stated that they volunteered after having heard favorable word-of-mouth reports of the earlier session. Beyond the fact that all teachers were volunteers, no constraint was placed on obtaining a sample of teachers. The teachers received no special rewards of recognition for their participation.

A total of seven leaders were involved in the four workshop sections. The leaders represented four groups commonly involved in the provision of staff development workshops. One series was taught by a consultant trained by Stallings in the EUTP. The remaining three series were taught by two person teams of leader-trainees. The other leader teams were compromised of a pair of local university education professors, a pair of local principals, and a team comprised of a local teacher and a half-time teacher/ half-time coordinator of local programs for gifted and talented students.

Because trainer team is an independent variable in this study, contrasts between the teams along four dimensions are discussed below. Data was gathered using a combination of Likert-like and open-ended items in the Leadership Experience Questionnaire (LEQ) (Schaffer, 1984). The LEQ was designed for the current study.

FORMAL EDUCATION AND YEARS OF EDUCATIONAL EXPERIENCE

The external consultant and the two university faculty members all had earned doctorates in Education and between 15 and 20 years of education-related experience. The principals both had masters degrees and between 20 and 30 years of experience. The full-time teacher had a masters degree and 6 years of teaching experience. Her fellow leader, the part-time teacher, part-time local coordinator of gifted programs, had a masters degree, an additional administrator's certification, and 15 years of educational experience.

PRIOR KNOWLEDGE OF TEACHER EFFECTIVENESS RESEARCH

The consultant had read widely in the research on teaching, as had one of the university professors. The second university professor had read and taught aspects of the research on teaching. One of the two principals reported having done prior reading on teacher effectiveness research; the other principal reported having read none of the research. The teacher/program coordinator had no prior

training in, and reported having done no prior reading of, research in the field. The teacher had participated in the EUTP workshops presented the previous year, but had done no additional reading on the topics covered in the workshops.

LEADERSHIP EXPERIENCE

On the dimension of experience in leadership and in training of educational professionals, the consultant and the university professors all reported having had many experiences as workshop leaders and course instructors, having served on several curriculum committees, and each having made over 10 presentations at conferences and conventions.

One of the principals estimated that he had led over 10 workshops and instructed over 10 courses; he stated that he had never served on a curriculum committee, presented at a conference, or served in a leadership role in a community organization. His fellow principal reported having never served previously as a workshop leader or as a presenter at a conference. He had served as an instructor in a college-level course on one prior occasion, and had served on five separate curriculum committees. He reported having over 10 leadership experiences in community organizations.

Both the teacher and the teacher/coordinator reported having served as workshop leaders 1-3 times previously, never having been a course instructor, and having presented at conferences 1-2 times.

GROUP PROCESS SKILLS

Responding to an eight-item subsection of the LEQ, the principals rated themselves as highly proficient in facilitating group processes. The program consultant, the university faculty members, and the teacher/coordinator team members rated themselves as possessing adequate though not exceptional group process skills. The teacher and teacher/coordinator both rated themselves as possessing very good social maintenance skills. In that one area they rated themselves clearly above the other groups.

PROCEDURE

Pre-workshop SOS data were gathered on the 27 teachers during January and February of 1982. Workshop series were held the following two months, and post-workshop data were gathered as soon after each group's fifth workshop meeting as was practical. In most cases, that meant that all post-workshop data were gathered within 10 days of the fifth workshop.

Although all 27 teachers completed the workshop series, problems with post-workshop data gathering and analysis eliminated several teachers from the final research data base. Analysis of pre-workshop SOS data, in which the six dropped cases were contrasted with the 21 (on some variables, 22) remaining cases, indicated no significant differences between those eliminated and those remaining. At least one case was lost from each group, with no group losing more than two cases.

Two hypothesis were tested. The first concerned replication of the basic Stallings findings. Stated in the null form, it was hypothesized that the teachers involved in the EUTP would not change their classroom behaviors after attending the workshops. The second concerned the generalizability of the EUTP across trainer groups. It was hypothesized that varying leader background/position would not significantly alter the level of behavior change of teachers involved in the workshops.

Preliminary analyses involved the entire group of teachers on the full 31-variable data set. A simple sign test was used to determine whether, on a

significant majority of the variables, teacher behavior changed in the desired directions. The full design called for two independent variables, group (consultant, university, principal, and teacher lead groups), and time of observation (Pre- and post-workshop, a repeated measure). Three aggregated SOS measures, all academic statements, interactive teaching, and organizing/off task activities, were dependent variables.

RESULTS

Data on the 31 specific SOS variables, pre- and post- EUTP are presented in Table 1. Mean change by the 21 teachers was in the desired direction on 24 of 31 SOS variables ($z = 2.87$, one-tailed $p < .001$). Overall, the teachers' altered their behavior patterns in the desired directions.

Table 1

Sample Means on 31 SOS Variables Before and After EUTP Workshops

Subscale Variable	EUTP Criterion	Pre-EUTP	Post-EUTP	Change
<i>Five Minute Interaction</i>				
All Academic Statements	80.00+	82.94	92.38	+9.44
Teacher Instructs/Explains	25.00+	21.72	27.26	+5.54
Teacher Asks Direct Questions	8.00+	13.03	14.03	+1.27
Teacher Asks Clarifying Questions	3.00+	.38	.56	+1.18
Teacher Calls on Different Students	6.00+	5.27	8.51	+3.24
Students Respond	8.00+	14.54	16.88	+2.34
Teacher Praises or Supports	4.00+	4.66	6.66	+2.00
Teacher Corrects	4.00+	2.98	3.12	+1.14
Teacher Corrects and Guides	2.00+	1.63	2.27	+1.64
Students Read Aloud	12.00+	2.33	4.56	+2.23
Teachers Read Aloud	10.00+	1.45	1.84	+1.39
All Organizing or Managing Statements	12.00-	13.72	5.29	-8.43
Teacher Working Alone	.00-	5.50	1.47	-4.03
Teacher Monitoring Written Work	5.00-	2.86	5.40	+2.54
All Behavior Statements	3.00-	2.68	1.63	-1.05
All Social Statements	2.00-	.32	.14	-.18
Intrusion	.00-	1.34	.57	-.77
Positive Interactions	2.00+	1.08	1.64	+1.56
Negative Interactions	.00-	.19	.14	-.05
<i>Classroom Snapshot</i>				
Reading Silently	15.00-	1.21	.60	-.61
Reading Aloud	6.00+	6.05	12.44	+6.39
Making Assignments	10.00-	4.21	3.96	-.24
Introduction/Explanation	25.00-	41.94	50.25	+8.61
Discussion/Review Assignments	10.00+	24.97	15.94	-9.03
Practice Drill	4.00+	1.50	1.21	-.29
Written Assignments	20.00+	4.22	8.26	+4.04
Taking Test/Quiz	5.00+	2.72	.91	-1.81
Social Interaction	.00-	.00	.30	-.30
Student Uninvolved	.00-	.60	.61	+1.01
Being Disciplined	.00-	1.80	.60	-1.20
Classroom Management	5.00-	8.46	3.35	-5.11

Table 2

Pre- and Post-workshops Rates of Teacher Activity in Three Stallings Observations System Aggregated Variables

Variable	Group*	Time			
		Pre-EUTP		Post-EUTP	
		-X	(SD)	-X	(SD)
All Academic Statements	C	81.85	(8.29)	89.50	(4.52)
	U	82.78	(17.19)	92.60	(4.56)
	T	74.60	(14.92)	92.20	(5.89)
	P	82.93	(8.02)	93.57	(3.78)
Interactive	C	63.27	(12.73)	71.55	(12.56)
	U	76.58	(18.24)	84.44	(12.41)
	T	78.56	(17.23)	80.00	(16.33)
	P	83.70	(13.27)	83.40	(12.45)
Organizing/Off-Task	C	19.92	(10.90)	13.25	(9.43)
	U	17.68	(19.11)	10.00	(9.19)
	T	22.60	(10.11)	8.00	(8.69)
	P	4.71	(3.22)	5.85	(7.30)

*C = Consultant; U = University; T = Teacher; P = Principal

Descriptive statistics for the four training groups on the three aggregated variables is presented in Table 2. Two features of the data are worthy of note before statistical analyses are presented. The majority of the teachers in the study exhibited behavior patterns at or above Stallings' criteria for effective teaching before the workshops began. Stallings' (1980) prior research had indicated that effective teachers spent 80% or more of their class time engaged in All Academic Statements.

The pre-workshop grand mean for the teachers in this study was 82% ALACS. Stallings' criterion for Interactive Teaching (INT) was 50%. The pre-workshop grand mean of this sample was 77%. Stallings recommended that teachers spend less than 20% of their time in Organizing and Off-Task activities. The teachers' pre-EUTP mean was 15%. It is possible that these scores were the result of a volunteer subject effect. (Rosenthal & Rosnow, 1975; Devlin-Scherer, 1984.)

A second observation concerned the drop in dispersion of scores from pre- to post-workshops. The entire sample's standard deviation on ALACS dropped from 12.5 pre-workshops to 4.5 post-workshops. The standard deviation for INT dropped from 16.1 to 13.3, and for ORGOF from 13.5 to 8.3. These data cannot be explained as a regression to a mean. On each variable the groups moved away from the national norms provided by Stallings.

As can be seen in Table 2 (above) and 3 (next page), teachers in the program spent significantly ($p = .001$) more time engaged in All Academic Statements after the workshops. The time by group interaction and main effect for group were not significant. All groups rose substantially on ALACS. Effects on Interactive Teaching were not significant over time or between groups.

Organizing/Off-task was the only aggregated variable on which groups differed significantly ($p < 0.47$). This significance may be partially due, however, to the reverse ceiling effect in the principals' group. The group, on average spent close to zero percent of their observed time organizing or off task. Thus, this contributed very little variance to the analysis. The effect of time approached

Table 3

Analyses of Variance: Group (4) by Time (2, repeated) for Three Aggregated SOS Variables

Variable	Source	All Academic Statements			
		<i>df</i>	SS	MS	F
All Academic Statements	Between Ss				
	Group (G)	3	295.48	98.49	.81
	S(G)	17	2069.24	121.72	
	Within Ss				
	Time (T)	1	1092.42	1092.42	19.26***
	G X T	3	194.55	64.85	1.14
	T X S(G)	17	964.23	56.72	
Interactive	Between Ss				
	Group (G)	3	1390.84	463.61	1.79
	S(G)	18	4668.30	259.35	
	Within Ss				
	Time (T)	1	165.56	165.56	1.00
	G X T	3	162.13	54.04	.32
	T X S(G)	18	2965.88	164.77	
Organizing/Off-Task	Between Ss				
	Group (G)	3	949.67	316.56	3.22
	S(G)	18	1770.72	98.37	
	Within Ss				
	Time (T)	1	431.63	431.63	3.54*
	G X T	3	371.71	123.90	1.02
	T X S(G)	18	2193.32	121.85	

* $p < .08$, ** $p < .05$, *** $p < .001$

significance ($p < 0.78$). Again, the analysis may have been blunted by a reverse ceiling effect. Many of the teachers were spending close to zero percent of their observed pre-workshop time organizing or off-task.

DISCUSSION

The hypothesis that the Stallings program could be replicated in a small town-to-rural context was largely confirmed. Teachers' classroom behavior rates on all of the 3 aggregated variables moved in the desired direction. For ALACS, the change was highly significant; for ORGOF it was marginally significant. Study-wide change was least for Interactive Teaching, though even there the movement was in the desired direction. While neither the interaction nor main effects for Interactive Teaching were significant, a visual examination of the data can provide interesting grounds for speculation. Both the principal and teacher-led groups exhibited essentially no group level change in behavior.

By contrast, the consultant and university professor-led groups gained on average of 9 and 10 percentage points respectively. Interactive Teaching involves a complex cluster of behaviors, and it is possible that the persons more familiar with the complexity of the literature and the reality of the concept did a better job of conveying that complexity to teachers and, eventually, were more

able to facilitate teacher change on this complex topic. Alternately it could be argued that, given that the groups led by principals and teachers scored so high on interactive teaching during the pre-observations, that they had no motivation to focus on those apparently already mastered skills.

The lack of dramatic effects in teacher behavior patterns, in spite of a relatively intensive inservice education program, may be attributed to at least two factors. The first is ceiling effects. On average, these teachers began the study performing at a very high level. A teacher who begins a EUTP spending over 80% of his/her time in ALACS and under 10% in ORGOF has a limited range in which to improve on those measures.

A second explanation may come from the effect of mid-year training. Emmer et al. (1981), in a classroom management change study, found that teachers trained early in the school year exhibited substantially more behavior change than teachers trained at mid-year. The workshops in the present study were all conducted in late winter when the school year was already more than half completed. Teacher behavior patterns may be more nearly set at that point, and large change would have been extremely difficult to achieve.

This study did not replicate the Anderson (1985) finding that an outside consultant was more effective at producing change than were local leaders. The consultant in the current study was the same person in the Washington DC implementation. It is possible that in the current study the consultant, having lived in the same state in which the workshops were conducted, was not sufficiently exotic to produce a strong experimenter effect. It is also plausible that the finding of the Washington study was an anomaly.

More broadly, the study provides specific, if limited, support for the generalizability of extended, teacher specific-change studies. Both the small town to rural context of the study and the fact that local implementors were not statistically inferior to their university-based or consulting colleagues at inducing teacher change suggest that cost-efficient, research-based, context-specific teacher change efforts can succeed. ■

REFERENCES

- 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, S. (1985). Teacher training techniques from four observational perspectives. *Journal of Classroom Interaction*, 20(1), 16-28.
- Brophy, J. and Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching (3rd ed.)*. New York: MacMillan.
- Bush, M. (1985). The complexity of institutionalising a program: Acquisition of training, observing, and computing capability. *Journal of Classroom Interaction*, 20(1), 6-15.
- Coladarci, T., & Gage, N. (1984). Effects of a minimal intervention on teacher behavior and student achievement. *American Educational Research Journal*, 21, 539-555.
- Crawford, J., Gage, N., Corno, L., Stayrook, N., Metman, A., Schunk, D., Stallings, J., Baskin, E., Harvey, P., Austen, D., Cronin, D., & Newman, R. (1978). *An experiment on teaching effectiveness and parent-assisted instruction in the third grade* (3 vols.). Stanford CA: Center for Educational Research at Stanford.
- Devlin-Scherer, W. (1984). *Implementation of a J.D.R.P. approved teacher effectiveness training program: Evaluating presage, process and long-term outcome*. Symposium conducted at the meeting of the American Educational Research Association, New Orleans.
- Emmer, E., Stanford, J., Evertson, C., Clements, B., & Martin, J. (1981). *The Classroom Management Improvement Study: An experiment in elementary school classrooms*. Austin: Research and Development Center for Teacher Education, University of Texas, 1981. (ERIC Document Reproduction Service No. ED 178 460).

- Gage, N. (1985). *Hard gains in the soft sciences: The case of pedagogy*. Bloomington, IN: Phi Delta Kappa.
- Good, T., & Grouws, D. (1979). The Missouri Mathematics Effectiveness Project: An experimental study in fourth grade classrooms. *Journal of Educational Psychology*, 71, 335-362.
- Goodlad, J. (1984). *A place called school*. New York: McGraw Hill.
- Rosenthal, R. (1976). *Experimenter effects in behavioral research* (2nd Ed.). New York: Irvington.
- Schaffer, G. (1984). *The Leadership Experience Questionnaire*. Charlotte, NC: University of North Carolina at Charlotte.
- Sirotnik, K. (1984). An inter-observer reliability study of a modified SRI observation system. *Journal of Classroom Interaction*, 19(1), 28-38.
- Stallings, J. (1980, December). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher*, 9(12), 11-16.
- Stallings, J. & Kaskowitz, D. (1974). *Follow through classroom observation evaluation (1972-1973)*. Menlo Park, CA: SRI International.
- Stallings, J., Needels, M., & Stayrook, N. (1978). *How to change the process of teaching basic reading skills in secondary schools: Phase II and Phase III, Final Report*. Menlo Park, CA: SRI International, 1978.
- Stringfield, S., Teddlie, C., & Suarez, S. (1985). Classroom interaction in effective and ineffective schools: Preliminary results from Phase III of the Louisiana School Effectiveness Study. *Journal of Classroom Interaction*, 20(2), 31-37.