RELIABILITY IN CLASSROOM OBSERVATION: OBSERVER EFFECTS AND STABILITY OF BEHAVIOR

By

MARJORIE RAGOSTA

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1974

ACKNOWLEDGMENTS

It would be difficult, indeed, to express adequate thanks for my indebtedness to Dr. Robert Soar. My association with his National Evaluation of Follow Through for more than three years was the most valuable experience in educational research that I can imagine. Dr. and Mrs. Soar provided me with training in the way I should go, and what I learned from them will remain with me throughout my research days. This dissertation is a direct result of the myriad opportunities opened to me by them.

Thanks are due, too, to the other members of my committee: Dr. John Newell, who always found time for me in his busy schedule, and Dr. Richard Anderson, who was my advisor in the Psychology Department. At times when I was bursting with questions and implications arising from learning theory or educational research, Dr. Newell's door would be standing open and he would be ready to listen, to argue, and to set me straight. While I was still in Psychology, Dr. Anderson gave me the freedom and encouragement to investigate coursework in the Foundations Department of the College of Education. Without his understanding I might not have made that change in my major which I have never since regretted making.

To each of my committee, many thanks.

To my friends at Educational Testing Service must go the credit for persistance in encouraging me to write. It would have been easy to neglect my dissertation in the crush of research activity, but my friends and colleagues would not let me. Educational Studies Division provided the help, incentive, time and continual support that I needed. My grateful thanks to all of my friends at ETS.

Data used in this study were collected under contract OEG-0-8-522394-3991(286) from the Follow Through Department, Division of Elementary and Secondary Education, Office of Education, to the Florida Educational Research and Development Council, Gainesville, Florida.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued...)

LIST OF TABLES

LIST OF TABLES (Continued...)

LIST OF TABLES (Continued...)

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Statement of the Problem

With the increased use of classroom observation systems, increased concern is evident over questions of (1) intra- and inter-observer reliability, (2) the instability of classroom behaviors over time, and (3) the effect of the presence of an observer in the classroom. Since reliability in the general sense refers to consistency across a series of measurements, the three concerns are intimately related and almost impossible to separate.

Test reliability is, itself, not a unitary concept. The standard error is used to measure reliability in terms of the variability of scores in repeated testing of the same individuals. Correlation is used to obtain "comparable-forms," "split-half," or "test-retest" reliability coefficients in an attempt to measure consistency in different kinds of testing situations (Cronbach, 1960). Analysis of variance has been shown to give a reliability measure equivalent to that obtained with correlation adjusted by the Spearman-Brown formula (Hoyt, 1955). Reliability always refers to consistency along a series of measurements but different reliability coefficients mean different things.

The situation becomes even more complex when the reliability of classroom observation systems is considered. Along with the reliability of the instrument itself as a measuring device, the reliability of the

recorders (observers) must be considered. Intra-observer reliability, i.e., the consistency with which an observer records his observations, as well as inter-observer reliability, i.e., the consistency of recording among different observers, must be considered.

As early as 1929, it was recognized that the largest component in the unreliability of observation systems is the instability of behavior from one observation to another (Barr, 1929). That is, in order to measure reliability, one must also measure the actual changes of behavior in the classroom, whether those changes are random or lawful, or due to the presence of an observer.

Since the efficacy of any specific classroom observation system must be in doubt until questions of reliability are resolved, this study was undertaken with the following objectives:

1. To estimate the intra- and inter-observer reliability of the Teacher Practices Observation Record, TPOR (Brown, 1968), and the Florida Climate and Control System, FLACCS (Soar, 1971),
2. To determine whether any of the behaviors measured by the TPOR and FLACCS factors changed significantly over time of day, and
3. To discover whether any changes in behavior over the day could be related to the observers' ratings of the effect of their own presence in the classroom.

Discussion of the Instruments

Teacher Practices Observation Record - TPOR

The TPOR was designed to measure the extent to which a teacher is, or is not, behaviorally in agreement with John Dewey's experimentalism. Each of the 31 pairs of odd and even items represent contrasting behaviors

that are traditional or experimental in nature. Although the TPOR was designed originally to obtain a single score measuring agreement with Dewey, in the Follow Through studies the items have been factor analyzed to produce factors representing several aspects of classroom behavior (Soar, 1973). A copy of the Teacher Practices Observation Record is included as Figure A-1 in Appendix A.

Florida Climate and Control System - FLACCS

FLACCS (Soar, Soar and Ragosta, 1971) developed out of the Florida Affective Categories System (Soar, 1971) which was a modification of the South Carolina Observation Record (Soar, 1966). FLACCS has two parts with 84 items and 78 items respectively. The first section, on classroom management, is concerned with the structure and control of the classroom by the teacher and aides, and with the responses of the pupils. The second section measures verbal and nonverbal expressions of positive and negative affect on the part of both teachers and pupils. A copy of FLACCS is included as Figure A-2 in Appendix A.

Discussion of the Concepts

The reliability of a classroom observation instrument is dependent on many things, not the least of which is the instrument itself. Test reliability is measured in terms of the whole test, or subtests depending on how scoring is done, but not generally in terms of individual items which would have less reliability. Similarly, each of the items within an observation instrument may separately have low reliability, but clusters or groups of items may achieve the reliability that none could claim alone. Data from FLACCS, and from the TPOR separately, were factor analyzed using principal axis factor extraction with multiple R

squared in the diagonal, followed by varimax rotation (Guertin and Bailey, 1970). T-scored items loading $\pm$ .40 or higher were pooled with unit weights to produce incomplete factor scores. Each factor purports to measure some aspect of classroom behavior, and the factor score might be expected to have greater reliability than an item score.

Test reliability is concerned with behavior -- the behavior demonstrated by the recorded responses to test questions. Reliability in classroom observation must be concerned not only with the behavior of the recorder (i.e. observer) but with the behavior of the teacher being recorded and the interaction between the observer and the teacher. That is, the reliability is dependent on consistency of recording between and within observers, the stability of the classroom behaviors of teachers and pupils, and the effect of the presence of an observer on the teacher and pupils.

Intra- and Inter-Observer Reliability

The primary function in the training of observers is to educate them to be consistent and identical recording instruments. To the extent that the observer is not consistent in the recording, reliability will be decreased. If liking or disliking a classroom affects the way the observer "sees" the behaviors in the classroom, the recording may reflect the observer's feelings rather than the "true" classroom behavior. To the extent that events occurring in the classroom color the way the observer records, reliability is decreased. In both of these examples a lack of consistency of recording within a single observer decreases intra-observer reliability. More obviously, to the extent that two or more observers seeing the same behaviors record them differently, the inter-observer reliability of the instrument is decreased.

Stability of Classroom Behavior

Granted that one could be sure of the reliability of items within a factor on an instrument and sure of reliability within and between observers, the reliability of an observation instrument also depends on the stability of the behavior in the classroom if the resulting score is used as "typical" of the classroom. A measure is reliable to the extent that the average difference between two measurements independently obtained in the same classroom is smaller than the average difference between two measurements obtained in different classrooms (Medley and Mitzel, 1963, p. 250). Using this definition, all reliability data from measures of teacher behavior over time will reflect real differences in teacher behavior from one observation to another.

Classroom behaviors in kindergarten, first and second grades include activities ranging from free play to very structured situations, from active game-playing to listening to stories, from show-and-tell to learning to read. Such variability of classroom situations could contribute to a lack of stability in the measurement of classroom behaviors.

Presence of Observers

It has already been stated that events in the classroom which affect the way the observer records diminish reliability. Additionally, to the extent that the presence of the observer differentially affects the behavior in various classrooms, reliability decreases. If one teacher is relatively unaffected by the presence of observers, while another who is apprehensive at first gradually relaxes, and a third becomes more and more upset, their differing responses to being observed may be reflected in the behaviors under observation, causing decreased reliability.

Significance for Educational Research

Knowledge about reliability relative to several groups of behaviors (i.e., factors) on two instruments measuring different aspects of classroom behavior has relevance beyond knowledge about the two instruments themselves. Quality control is as necessary in research as in industry, and awareness of influences tending to decrease reliability can lead to methods of overcoming the deficiency.

Knowledge of intra- and inter-observer reliability may change methods of training. Knowing the stability of different kinds of behaviors and the effect an observer might have on the classroom has implications for deciding the length of time to be spent in a classroom visit. What we learn from a closer look at the various aspects of reliability may hopefully help to make the observation systems better, the training of observers better, or the data processing more exact, while increasing our awareness of classroom observation as an interactive process.

CHAPTER II

REVIEW OF THE LITERATURE

Classroom Observation

Classroom observation is not a new procedure. What is relatively new is an attempt to measure the teaching process through systematic observation techniques and to relate those process measures to the growth of students. Morsh and Wilder (1954) found that attempts to validate process criteria by correlating them with measured pupil growth had been, up to that time, generally unsuccessful. The development of Flanders' Interaction Analysis opened a new avenue of research (Flanders, 1965, 1967, 1970). In the proliferation of observation systems since that time, the single instrument that has been most often used as the basis for an expanded system or as part of the synthesis of old systems into new ones is Flanders' (Rosenshine and Furst, 1973, p. 143). The Observation Schedule and Record (Medley and Mitzel, 1958, b) also led to large amounts of process-product research.

Government support for the evaluation of large scale education programs, such as Head Start and Follow Through, has also encouraged the use of systematic observational techniques. Assessment of program differences in the Planned Variation phase of Head Start and Follow Through has been accompanied by attempts, not only to collect observation data, but in some cases, to relate observation data to the growth of pupils. Systematic observation instruments which have been used in Head Start

and Follow Through evaluation include the Personal Record of School Experience - PROSE (Medley, 1968), The Teacher Behavior Survey Instrument and the Child Behavior Survey Instrument (Katz, 1968; reported also by Gordon, 1973), Stanford Research Institute's Classroom Observation Instrument - COI (Stallings, 1972), the Teacher Practices Observation Record - TPOR (Brown, 1968; Soar, 1971, 1973), the Reciprocal Categories System - RCS (Ober, et al., 1968; Soar, 1971, 1973), the K-1 version of the Taxonomy of Cognitive Behavior (Ragosta, et al., 1971; Soar, 1971, 1973), and the Florida Climate and Control System - FLACCS (Soar, 1971, 1973).

Much of the data from Head Start and Follow Through are not yet generally available. One can only hope that the research involving observation systems might be further explicated and the best of the systems used to replicate and add to the findings.

> The orderly growth of knowledge might be promoted if researchers were less ready to devise their own observation systems and instead worked more closely within existing systems. (Neujahr, 1972, p. 228)

Validity and Reliability

Validity

Dopyera and Lav (1969), in surveying the procedures for assessing the program environments of Head Start and other pre-school children raise, among other issues, the metric problems of reliability and validity. The validity advocated in the addendum to their final report is predictive validity (i.e., association of a behavioral measure with independent child-change measures). Although some would oppose this position as being premature (Jackson, 1967), support for this definition of validity is high. Johnson and Bolstad (1973), in writing of the

validity of naturalistic behavioral data, stress the need for establishing predictive validity. Gage (1967) considers an educational study a bit of research only when category development and use is accompanied by achievement measures. Rosenshine (1971) reports on approximately 30 studies in which data from systematic observation had been related to student achievement or growth. Indeed, Rosenshine and Furst (1973) advocate an emphasis on predictive validity before reliability.

> The predictive validity may diminish or increase as soon as reliable observational data are used in a statistical analysis, but at least the researcher would know whether the variables are worth studying. (Rosenshine and Furst, 1973, p. 169)

Many developers of observation systems overemphasize observer agreement at the expense of validity. Medley and Mitzel recognized this as early as 1963 when they noted that concern with observer agreement may have eliminated valid items from an observation instrument (Medley and Mitzel, 1963, p. 263 ff). Additional evidence in this regard comes from researchers in behavior modification.

> Finally, high reliability bears no relation to the validity of a particular measure. It seems that reliability can be purchased at the expense of the substantive meaning of a particular measurement. (O'Leary and Kent, 1973, p. 88)

Reliability

Different kinds of reliability measures mean different things. In test reliability both the standard error and correlation have been advocated as two differing approaches to the problem of reliability (Thorndike, 1950). Correlation has been used to obtain coefficients of stability, or coefficients of equivalence and stability depending on the kind of test data being correlated (Cronbach, 1960). Analysis of variance has been shown to give the same reliability measure as correlation adjusted by the Spearman-Brown formula (Hoyt, 1955).

With classroom observational measures reliability becomes more complex. It has often been reported in terms of agreement between observers (Weick, 1968; Flanders, 1965, 1967). In the Head Start and Follow Through studies mentioned earlier, percent agreement was used by both Katz and Stallings. In one case the formula used was:

$$\frac{\text{\# of agreements}}{\text{\# of agreements + \# of disagreements}}$$

where non-occurrence of a behavior was included in the calculations (Katz, 1968). Two trained observers in that study were able to achieve an 84.6 percent overall percentage of agreement. Stallings (1972) reports a mean percent agreement of 86.8 percent for variables least frequently recorded and 69.0 percent for variables most frequently reported, based on eight pairings of observers in different locations.

The question of when percent of agreement fitures are "good" or "good enough" is a difficult question to answer. There are no established criteria and very little data by which a relative evaluation could be made. Johnson and Bolstad (1973) speak of 80-85 percent agreement as traditionally computed as a realistic upper limit for their complex coding in naturalistic observation.

Reliability figures defined in terms of percent agreement have generally been based on the total data collected rather than the data as used in further analyses.

> The reliability figures reported [for Flanders' Interaction Analysis] are based on total data matrices and do not represent interrater reliability for subscores and subsections of the matrix that are employed in recommended data analyses. (Mitchell, 1969, p. 710)

Johnson and Bolstad (1973) acknowledge percent agreement as a basis for one kind of reliability but propose one central prescription - the agreement data should be computed on the score used as the dependent variable. They see no sense in reporting overall average agreement data except perhaps as a bow to tradition. Confirmation for this point of view comes from data out of the National Evaluation of Follow Through.

> The data do illustrate the inadequacy of presenting reliability in terms of observer agreement for a total instrument which is common. The RCS data, for example, indicate that reliabilities of separate measures ranged from indicating doubtful value to ones that were quite high. To cite one overall value for the instrument as a whole does not seem to be useful. (Soar, 1973, p. 129)

In addition to percent agreement, correlation has been used in Head Start and Follow Through observational data to report reliabilities. In 1968 a pilot study was conducted using a preliminary form of PROSE (Medley, 1968). The reliabilities for all 67 alternatives were reported by correlating the total number of times an item was checked by one observer with the total number of times it was checked by the other observer in four visits to classrooms involving 27 pupils. The correlation was stepped up by the Spearman-Brown formula to obtain the reliability of the total observation time. Many of the reliabilities were significant; some were close to zero. Taken as a whole, Medley felt the data clearly indicated that PROSE records made by different recorders at different times detected stable differences between pupils and teachers. However, Medley recommended that reliability of scales -- not items -- be used in the Head Start Longitudinal study itself.

Reliabilities for FLACCS and TPOR were calculated from limited data (Soar, 1973). The 16 observers who collected the actual classroom data

observed a video tape of one teacher before beginning their winter observations and a tape of another teacher at the end of their field work. Data from these observations were used to estimate reliability by Hoyt's (1955) analysis of variance procedure. In his discussion, Soar points out that reliabilities of some factors may have been overestimated while other estimates were meaningless because of the limitations of the use of video tape.

All of the measures of reliability mentioned above are concerned with inter-observer (between-observer) reliability, i.e., whether all observers are equivalent recording instruments. Brown, et al., (1968) was more concerned with intra-observer (within observer) reliability, i.e., whether an observer is consistent over time in recording his observations. In a lucid article discussing reliability, he reports both inter- and intra-observer reliabilities for untrained observers using the TPOR.

The inadequacy of most of the reported measures as indices of reliability has been discussed widely (Medley and Mitzel, 1963; Westbury, 1967; Dopyera and Lay, 1969; McGaw, et al., 1972; Soar, 1973). Indeed, Medley and Mitzel have distinguished three coefficients: the reliability coefficient, based on observations made by different observers at different times; the coefficient of observer agreement, based on observations made by different observers at the same time; and the stability coefficient based on observations by the same observer at different times. They place less importance on observer agreement and more importance on the reliability coefficient which best estimates the accuracy of measurements. They find it wasteful to send two observers into the classroom to measure the same behavior, preferring to have two estimates of behavior by using the observers at two different times. Westbury (1967),

reporting a lack of stability in teacher interaction patterns, feels that more is needed for instrument reliability than a measure of observer agreement. Others agree:

> Establishing that two or more observers agree a large portion of the time is a reasonable approach when that figure is an unbiased estimate of the overt occurrence of phenomena. Unfortunately, the fact is that many, if not most, reliability estimates presented in this fashion are not, in fact, unbiased...(Dopyera and Lay, 1969, p. 51)

O'Leary and Kent (1973) discuss the reliability of observers using O'Leary's nine category observational code for measuring the disruptive behavior of children in a classroom setting. They report higher reliabilities when the trainer was absent than when he was present during observations and higher reliabilities as computed by the observers than reliabilities computed by the experimenter. Several studies report observer bias - a drift in recording to match those with whom one is paired (O'Leary and Kent, 1973; Skindrud, 1973). Such data suggest that individual observers or groups of observers should not be confounded with different experimental treatments but should be distributed across treatments.

In view of the many problems associated with the concept of reliability, particular attention must be paid to the usefulness of the types of reliability reported in observational data. The timing of reliability data is also an important factor. Many studies report reliabilities on data collected <u>before</u> the instrument is actually used in a study. Such reliability data is usually obtained during the training of observers. Reliabilities on FLACCS and TPOR as well as PROSE fall into this category although in the case of FLACCS and TPOR additional reliability data was collected <u>after</u> the major data collection effort

(Soar, 1973; Medley, 1968). Reliability data is sometimes obtained during the actual data collection. The COI reliability data were collected by having a pair of observers make simultaneous recordings in eight classrooms during the actual data collection effort (Stallings, 1972). Medley and Mitzel (1963) approve the measurement of reliability by an analysis of variance in the actual data analysis phase. Dopyera and Lay (1969) reemphasize that approach as do McGaw, Wardrop, and Bunda (1972) although, to date, no such reliabilities have been reported for observational data in Head Start or Follow Through.

Stability of Classroom Behavior

As early as 1929, it was believed that the instability of teacher behavior from one lesson to another was the major component in the unreliability of observation systems (Barr, 1929). But findings have been mixed, with some behaviors seemingly stable, others not stable.

Jersild (1939) used an observation scheme tallying pupil behaviors in seven categories: cooperative, critical, leadership, work-spirit, experimental, recitational, and self-initiating. Pairs of observers visited teachers ten times for half an hour each visit. Stability of behavior was measured by correlating the sum of the first five half-hour observations of each teacher with the sum of the last five observations for each category of behavior, adjusted by the Spearman-Brown formula. The first four categories mentioned above had stability coefficients of .88, .82, .90, and .94, respectively, although the percent agreement between pairs of observers for the same data were 70, 70, 76, and 43. Lower stability coefficients (.44, .68, and .60, respectively) were reported for the last three categories although their percents of agreement

were higher (87, 83, and 77, respectively). Thus, student behaviors categorized by Jersild as cooperative, critical, leadership and work-spirit appear to be relatively stable over time while those categorized as experimental, recitationa, or self-initiated are less stable.

Hughes (1959), on the basis of a small study involving six teachers who were observed during four half-hour sessions, concluded that teacher behavior patterns are stable over time. Her data were based upon an average of 250 agreed-upon teacher acts during each half-hour observation period. All agreed-upon acts were tallied as frequencies in seven categories: controlling functions, imposition, facilitating, content development, personal response, positive affectivity, and negative affectivity. Hughes felt that the number of controlling acts exhibited by a given teacher in different situations does not vary significantly when compared with the interaction between situations and teachers.

Soar (1973) reports on the stability of behavior in 17 Follow Through classrooms observed for one day each (12 five-minute time samples during the day) during the first week of school in September, again in late October or early November, and finally again during January, February, or March. There were no significant differences across time for six of the nine factors of the Florida Climate and Control System, FLACCS, or for any of the seven factors of the Teacher Practices Observation Record, TPOR. The exceptions were interesting. FLACCS 1, Strong Control, had means of 52.1, 49.8, and 48.5 for the three time periods while Factor 7, Pupil Negative Affect, had means of 54.8, 53.2, and 49.7. Both factors had significant differences over time. One other factor demonstrated a significant difference. Factor 6, Work Without Teacher had means of 49.3, 53.1, and 50.7 and was statistically significant.

Bellack, et al., (1966) reports an astonishing degree of stability over time in the pattern of discourse both for teachers and for pupils.

Mitzel and Rabinowitz (1953) and Medley and Mitzel (1958a) on the other hand, found that the variation in teacher verbal behavior from observation to observation provided a major source of variability. Observers in that study were self trained on Withall's technique (Withall, 1949) categorizing all teacher statements as (1) learner-supportive, (2) accepting or clarifying, (3) problem structuring, (4) neutral, (5) directive, (6) reproving or disapproving, or (7) teacher-supportive. Categorization in this technique is based not on the content of the statement but on the intent of the teacher making it. The index used in this study was the sum of statements in categories 1 - 3 (learner-centered statements) divided by the total number of statements for each of four teachers by each of two observers on each of eight visits.

> The significant effect for teacher-by-visit is an answer to the questions posed in this study concerning differences among teachers, and stability of teachers' verbal behavior from time to time. It was concluded that teachers showed marked fluctuations in their verbal behavior; "real" differences among teachers were apparent in spite of these fluctuations. (Mitzel and Rabinowitz, 1953, p. 15)

McGaw, et al., (1972) state that variability of the object of observation is the most important source of error variance, and unless stable estimates can be obtained variability across classrooms or teachers will inevitably be swamped by variability within. They take issue with the Medley and Mitzel (1963) position that instability of teacher behavior is due to random error in the person, environment or both.

> This implies that the characteristic being measured is stable in a sense that does not allow of lawful change. While this may be a reasonable assumption in relation to relatively enduring aspects of personality, it is unreasonable when other types of behavior patterns are being observed. (McGaw, Wardrop, and Bunda, 1972, p. 16)

A lack of knowledge and need for more research on the stability of teacher behavior has been pointed out repeatedly (Rosenshine, 1970; McGaw, et al., 1972).

Observer Effects

Medley and Mitzel (1963) point out that one of the reasons given for failing to include classroom observation in research studies is that the presence of an observer in the classroom is so disturbing that the behavior observed cannot be regarded as typical. The authors state that the objection has no completely satisfactory answer but they agree that to know how teachers and pupils behave during observation seems better than to know nothing at all about how they behave.

Mitzel and Rabinowitz (1953) in their study described earlier felt that teacher verbal behavior was markedly influenced by the presence of observers in the classroom, particularly during the early visits. In addition to the impressions of the observers that initially they were watching "tailor made" lessons, they use logic to support this conclusion.

> On logical grounds, it appears likely that the net effect of "preparation for being observed" is to reduce the differences among teachers in all statement categories and to have a general depressing effect upon the total variance for each category. Though it cannot be proved conclusively, this is a plausible explanation for the increase in the among-teachers mean square and in the total variances from the first series of visits to the second. (Mitzel and Rabinowitz, 1953, p. 14)

Masling and Stern (1969) had seven observers trained in the use of rating scales. Twenty-three teachers were observed for a period of two days by a single observer. An average of 33 five-minute observations were made over the two-day period. They assumed that if the hypothesis that the effect of the presence of the observer diminishes gradually over time was correct, then correlations between the earliest observations and the later ones should be lower than the correlations between observations at a later period. Data were reported for correlations between observations 1-3 versus 7-33, 4-6 versus 7-33, and 7-9 versus 10-33. Of the ten ratings made, only two, "class climate-freedom" and "class climate-tension," produced results to support the hypothesis. There was no consistent discernible pattern over time. Two alternative conclusions were presented: (a) observer influence is negligible; the lack of consistency over time was found because the variables under study occur episodically; or (b) the effects of the observer are more complex than has been foreseen and affect various aspects of teacher and pupil behavior differentially.

Samph (1968) made tape recordings in classrooms both without the teacher's knowledge and with the teacher's knowledge. Following the tape recording, observers collected live data using Flanders' Interaction Analysis. Data were given on five variables: amount of praise, amount of criticism, total teacher acceptance of pupil's ideas, the I/D ratio (number of indirect/direct teacher behaviors) and the i/d ratio (similar to the I/D but omitting the primarily substantive categories). For only the first two variables were significant differences found between observations made with and without an observer present in the classroom.

With the presence of the observer the amount of praise increased and the amount of criticism decreased significantly. None of the other variables showed significant change.

The two studies just mentioned illustrate the two basic paradigms by which reactive effects have been studied: by comparison of the effects of various levels of obtrusiveness of observations or by the study of behavioral stability over time. Johnson and Bolstad (1973) review the literature on reactivity in naturalistic observation and on the basis of these studies discuss four sources of interference: (1) the conspicuousness of the observer, (2) the individual differences of the subjects, (3) the personal attributes of the observer and (4) the amount of rationale given subjects for being observed. They feel there can be little doubt that the entire question of reactivity has been inadequately researched and raise the point that not only may reactivity seriously limit the generalizability of naturalistic observation data but it may interact with the variables under study to confound results.

CHAPTER III

PROCEDURE

Sample

President Johnson initiated the Follow Through program in 1968 to capitalize upon and supplement the gains made by disadvantaged youngsters in pre-school programs such as Head Start. As in Head Start, Follow Through took advantage of planned variation; that is, local communities could elect to be self-sponsored or could choose to be sponsored by one of a number of program models (Maccoby and Zellner, 1970). The models differed widely in philosophy. The Becker-Englemann model and the token economy of Bushell used behavior modification techniques. Gordon's parent education model focused on parent involvement in the motivation and education of their children. Nimnicht stressed the responsive environment and autotelic learning while Gotkin emphasized matrix games. Educational Development Center (EDC) sponsored an open education model based on the British Infant School. The Bank Street College of Education approach and the Tucson early education model completed the group of planned variation sponsors whose classrooms were observed in the present study.

Data were collected in 289 classrooms representing the eight Follow Through sponsors described above and a number of comparison (non-Follow Through) classrooms. Three grade levels were represented: kindergarten; first grade, divided into an entering-first (i.e., a first grade without a kindergarten) and continuing-first; and second grade. A summary of

the distribution of classrooms by sponsor and by grade level is given in Table 1.

TABLE 1

DISTRIBUTION OF CLASSROOM BY SPONSOR AND BY GRADE LEVEL

	Grade Levels				
Sponsors	K	E-1	C-1	2	Total
Bank Street	9	8	10	7	34
Becker-Englemann	8	9	10	4	31
E.D.C.	8	10	6	4	28
Gotkin	12	--	8	--	20
Nimnicht	11	8	11	7	37
Parent Education	8	10	9	6	33
Tucson	8	4	9	8	29
Bushell	7	5	10	--	22
Comparison	15	13	18	9	55
TOTAL	86	67	91	45	289

Data Collection

Systematic classroom observations were made during January, February and March of 1971, using the Florida Climate and Control System (FLACCS) and the Teacher Practices Observation Record (TPOR). Fourteen graduate students had been hired in the summer of 1970 to be observers during the winter of 1971. These observers were trained in a scheduled credited course during the fall quarter at the University of Florida. These 14 observers plus four permanent members of the evaluation team collected

the observation data. Team memberships were changed periodically during the ten weeks of observation and within the constraints of scheduling each observer was assigned to as many sponsors' classrooms as possible. The number of classrooms observed each week by sponsor is given in Table 2.

Both FLACCS and the TPOR instruments were set up so that each completed instrument contained a series of three five-minute observations:

Series 1	Series 2	Series 3	Series 4
Observation	Observation	Observation	Observation
1 2 3	4 5 6	7 8 9	10 11 12

During a total day in the classroom, four series of observations were made using each of the instruments, two series by one observer and two by the other.

The rate of observing was dependent on the length of the school day. If a kindergarten class was three hours long, the rate for that classroom would be four five-minute observations per hour. The rate for a six-hour class would be two observations per hour. Each series, therefore, represented one-fourth of a complete school day, however long that day might be.

At the end of the day, in the classroom, each of the two observers completed ratings on (1) classroom attitude toward observers, and (2) attention to observers:

(1) Classroom attitude toward observers:

Very Hostile	Up-tight	Neutral	Polite	Friendly, Open
1	2	3	4	5

TABLE 2

NUMBER OF FOLLOW THROUGH AND COMPARISON CLASSROOMS
OBSERVED EACH WEEK DURING A TEN-WEEK PERIOD BY SPONSOR

	Week										
Sponsors	1	2	3	4	5	6	7	8	9	10	Total
Bank Street	9	12	3	2	8						34
Becker-Englemann			6	2	4	11	5		3		31
E.D.C.	10	8	2	2		1		3	2		28
Gotkin				4	2					14	20
Nimnicht		3	10	7	2				6		37
Parent Education	9			4		5	9	2	3	1	33
Tucson		3	10	7	2				7		29
Bushell				4	2	5		10		1	22
Comparison	2	5	6	7	4	4	5	9	9	4	55
TOTAL	30	31	27	35	24	26	29	37	30	20	289

(2) Are the observers in the classroom:

A Focus of Attention		A Normal Disruption		Ignored
1	2	3	4	5

Data Analysis

In the original data analysis, tallies for each item were summed across the four series (12 observations) for each teacher. Means and standard deviations for each item were calculated across all teachers. Items with very small means and/or variances were then either eliminated or pooled with related items. An area transformation was then carried out, item by item, to make the data as nearly normally distributed as possible. The data for each observation system were then factor analyzed separately using principal axis factor extraction with multiple R squared in the diagonal, followed by varimax rotation. Nine factors were identified for FLACCS and seven for the TPOR. The composition of the factors including item description and loadings are presented in Tables B-1 through B-16 in Appendix B. Incomplete factor scores were calculated for each of the teachers on each of the factors by pooling the T-scores from the area transformation for those measures which loaded $\pm$ .40 or above. The incomplete factor scores of the nine FLACCS factors and the seven TPOR factors for each of the 289 teachers in the study were then used in further analyses to differentiate teacher and pupil behaviors across the eight Follow Through programs (and comparison classrooms) and to relate those classroom behaviors to the mean academic growth of pupils (Soar, 1973).

For the present study the identical procedure of obtaining incomplete factor scores was followed, except that, instead of summing the data for each classroom across all 12 observations (four series of three observations), the sums for each of the four series were used. In effect this gave four scores for each item for each classroom instead of one. Each score represented three-five-minute time slices of behavior over one-quarter of the school day. The data were T-scored by area transformation and incomplete factor scores calculated. The end result was four repeated-measures for each teacher on each of the nine factors of FLACCS and the seven factors of the TPOR.

Reliability

Incomplete factor scores over four periods of the day for 289 teachers were analyzed using a one-way repeated measures analysis of variance. An overall reliability score for each factor was computed using Hoyt's formula (4) (1955, p. 52).

$$r_{11} = \frac{MS_I - MS_e}{MS_I}$$

where MS_I and MS_e designate the mean square for individuals and for error respectively. All four factor scores for each teacher entered into the overall reliability computation for each factor.

Two scores for each teacher based on data collected by the same observer were analyzed using a repeated-measures analysis of variance. The score calculated from formula (4) using only these data represents within-observer reliability. Since there were two observers in the classroom, each recording two series of observations, two such intra-observer reliabilities were calculated.

A similar analysis of variance was used to analyze the data of series 2 and series 3 done by different observers. Using formula (4) again, an inter-observer reliability score was calculated.

Intra-observer and inter-observer reliabilities were calculated using only two of the four time periods for each classroom. Each of these reliability figures was adjusted by the Spearman-Brown formula (Cronbach, 1960, p. 131) for obtaining the reliability expected from twice as much data.

Stability of Classroom Behavior

The repeated-measures analyses of variance using all four incomplete factor scores for each classroom provided data on the stability of classroom behaviors for each of the nine factors of FLACCS and each of the seven factors of the TPOR.

Observer Effects

The ratings of classroom attitude and attention to observers, made by each of the observers at the end of each day of observation, were summed to produce one score for each classroom on each rating scale. Classrooms were blocked on the attitude and attention ratings and the incomplete factor scores for each FLACCS and TPOR factor were subjected to a two-way repeated measures analysis of variance. This was done to determine whether perceived differences in classroom attitude or attention to observers were accompanied by any significant differences in classroom behaviors or by any significant interaction effects.

CHAPTER IV

RESULTS

Incomplete factor scores for nine FLACCS factors and seven TPOR factors were generated for 289 Follow Through and non-Follow Through classrooms over four periods of time-of-day. The factor scores were used in further analyses in three related substudies investigating (1) the reliability of the observation instruments, (2) the stability of classroom behavior, and (3) the effect of the observer in the classroom. Results are reported separately for each of these areas of investigation.

Reliability

Reliabilities derived from several repeated measures analyses of variance are given in Table 3 for the nine factors from the Florida Climate and Category System and in Table 4 for the seven factors of the Teacher Practices Observation Record.

Overall reliabilities are based on four series of observations (four repeated measures) for each of the 289 teachers in the study. The within-observer or intra-observer reliabilities are based on the two series of observations (two repeated measures) done by the same observer. Within-observer reliabilities are broken into those collected early in the day by the first observer and those collected later in the day by the second observer in each of the 289 classrooms. The between-observer or inter-observer reliabilities are based on the middle two series of observations

TABLE 3

OVERALL, INTRA-OBSERVER AND INTER-OBSERVER RELIABILITIES FOR NINE FLACCS FACTORS*

	FLACCS Factor								
Reliability	1	2	3	4	5	6	7	8	9
OVERALL	.84	.81	.78	.74	.69	.70	.83	.64	.84
Intra (early)	.74 (.85)	.74 (.85)	.82 (.90)	.79 (.88)	.81 (.90)	.62 (.77)	.71 (.83)	.77 (.87)	.86 (.93)
Intra (late)	.84 (.91)	.69 (.82)	.83 (.91)	.67 (.80)	.72 (.84)	.65 (.79)	.78 (.88)	.73 (.85)	.81 (.90)
Inter	.69 (.82)	.70 (.82)	.57 (.73)	.49 (.66)	.43 (.60)	.57 (.73)	.65 (.79)	.38 (.55)	.70 (.82)

*Figures in parentheses have been adjusted by the Spearman-Brown formula to produce expected reliabilities for four series of observations rather than two.

TABLE 4

OVERALL, INTRA-OBSERVER AND INTER-OBSERVER RELIABILITIES FOR SEVEN TPOR FACTORS*

	TPOR Factor						
Reliability	1	2	3	4	5	6	7
OVERALL	.78	.71	.75	.62	.80	.68	.82
Intra (early)	.80 (.89)	.78 (.88)	.80 (.89)	.61 (.76)	.77 (.87)	.79 (.88)	.82 (.90)
Intra (late)	.68 (.81)	.77 (.87)	.74 (.85)	.62 (.77)	.71 (.83)	.74 (.85)	.75 (.86)
Inter	.65 (.79)	.45 (.62)	.56 (.72)	.41 (.58)	.65 (.79)	.37 (.54)	.68 (.81)

*Figures in parentheses have been adjusted by the Spearman-Brown formula to produce expected reliabilities for four series of observations rather than two.

(two repeated measures) done by two different observers in each of the 289 classrooms.

Each intra-observer and inter-observer reliability figure is equivalent to the correlation between the two observations adjusted by the Spearman-Brown formula (Cronbach, 1960, p. 131) for obtaining the reliability coefficient of the two series of observations together. Since each intra-observer and inter-observer reliability figure as presented was based on only two of the four series of observations, each of those figures was again adjusted by the Spearman-Brown formula to obtain the reliability expected from twice as much data. Adjusted reliability scores are presented in Tables 3 and 4 in parentheses.

Unadjusted intra-observer reliabilities range from coefficients of .62 to .86 for FLACCS factors and from .61 to .82 for TPOR factors. Scores adjusted by the Spearman-Brown formula are higher, ranging from .77 to .93 for FLACCS factors and .76 to .90 for TPOR factors. The intra-observer reliabilities measure the consistency of recordings within observers.

Unadjusted inter-observer reliabilities range from coefficients of .38 to .70 for FLACCS factors and from .37 to .68 for TPOR factors. Scores adjusted by the Spearman-Brown formula range from .55 to .82 for FLACCS factors and from .54 to .81 for TPOR factors. The inter-observer reliabilities measure the consistency of recordings across observers.

Overall reliabilities range from .64 to .84 on the FLACCS factors and from .62 to .82 on the TPOR factors.

Stability of Classroom Behavior

Florida Climate and Control System (FLACCS)

Mean factor scores on nine FLACCS factors over four periods of the day for the 289 Follow Through classrooms are presented in Table 5. Also included are the F's from a one-way repeated-measures analysis of variance of the data for each factor. Graphs of the data appear as Figure 1.

Significant differences in classroom behavior across four periods of the day characterize all FLACCS factors except for Factor 9, Teacher Positive Affect. Factor 1, consisting of behaviors representing strong control and negative affect on the part of the teacher, shows a significant increase over time along with the behaviors of Factor 7, Pupil Negative Affect. Classrooms also tend to move toward more free choice in the later part of the day (Factor 2). The last half of the day tends to show less Teacher-Pupil Supportive Behavior (Factor 3), Gentle Control (Factor 5), Work Without the Teacher (Factor 6), and Teacher Attention in a Task Setting (Factor 8). In general the data support the belief that most of the "work" of the classrooms tends to be done earlier in the day.

Teacher Practices Observation Record (TPOR)

Mean factor scores over four periods of the day for the 289 Follow Through classrooms are presented in Table 6 along with the F from a one-way repeated-measures analysis of variance for each of the seven TPOR factors. Graphs of the data appear as Figure 2.

There were highly significant changes over the day for three TPOR factors: Factor 1 - Convergent Teaching, Factor 4 - Undifferentiated

TABLE 5

MEAN FACTOR SCORES FOR 289 CLASSROOMS OVER FOUR PERIODS OF TIME OF DAY FOR EACH OF NINE FLACCS FACTORS

FLACCS	Name	Time of Day 1	2	3	4	F
1	Strong Control	49.28	50.64	50.84	50.97	19.23**
2	Pupil Free Choice vs. No. Choice	49.55	50.46	50.21	50.62	8.11**
3	Teacher-Pupil Supportive Behavior	50.31	50.73	49.92	49.58	7.27**
4	Non-verbal Gentle Control	49.77	50.63	49.66	50.05	2.71*
5	Gentle Control	50.74	50.65	49.86	49.50	6.06**
6	Work Without Teacher	50.13	50.85	50.25	49.60	4.48**
7	Pupil Negative Affect	49.44	50.53	50.68	50.85	14.10**
8	Teacher Attention	49.63	50.74	49.87	49.84	4.34**
9	Teacher Positive Affect	49.64	49.62	49.59	49.59	.93

* $p < .05$ ** $p < .01$

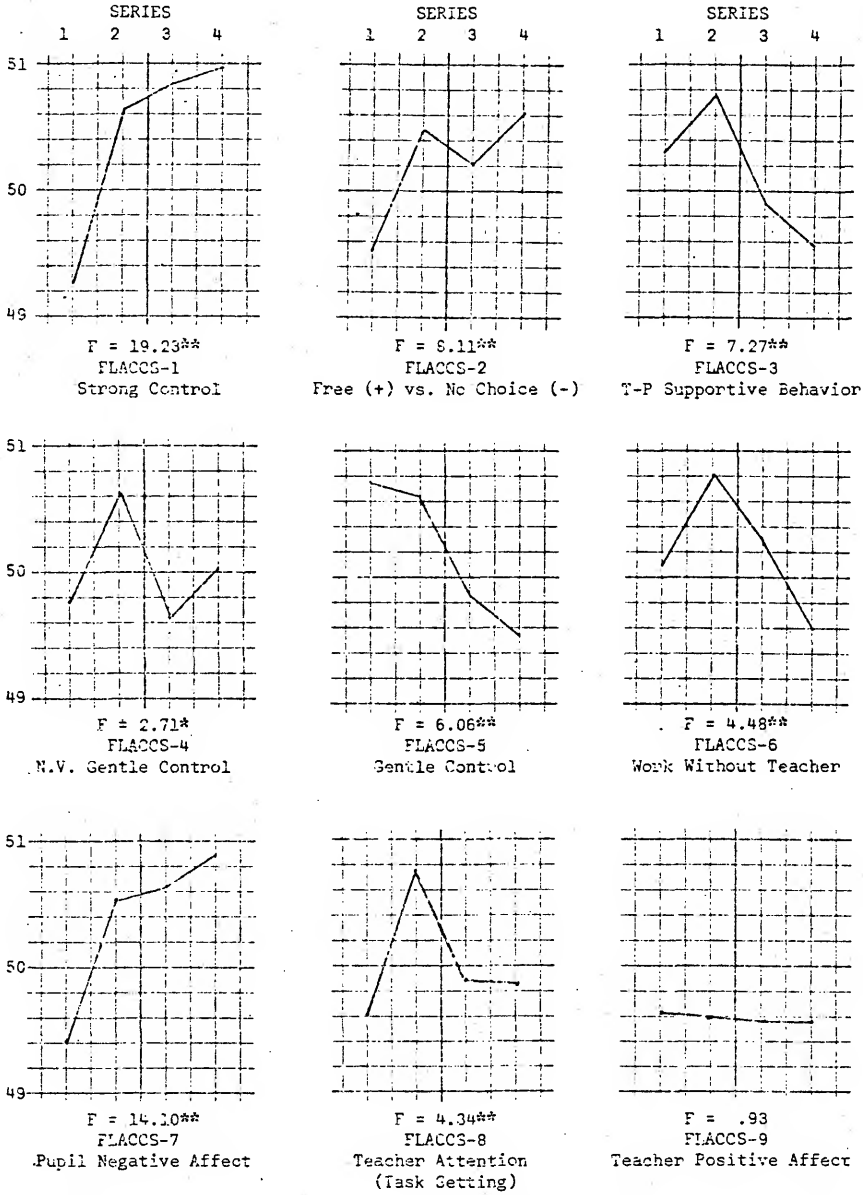

FIGURE 1. Graphs of mean factor scores for 289 classrooms across four series of observations during the day on each of nine FLACCS factors.

TABLE 6

MEAN FACTOR SCORES FOR 289 CLASSROOMS OVER FOUR PERIODS OF TIME OF DAY FOR EACH OF SEVEN TPOR FACTORS

FLACCS	Name	Time of Day 1	2	3	4	F
1	Convergent Teaching	50.46	50.77	49.76	48.58	17.79**
2	Experimental Teaching	50.48	50.52	50.68	50.32	2.03
3	Teacher Discourages Exploration	50.18	50.52	50.69	50.38	2.68*
4	Undifferentiated Teaching	50.40	49.18	50.68	51.60	9.47**
5	Pupil Free Choice vs. Teacher Structured Activity	49.87	50.19	50.44	51.08	7.56**
6	Unnamed	49.72	50.02	50.36	49.77	2.35
7	Exploration of Ideas vs. Textbook Teaching	50.46	50.12	50.36	50.51	1.91

* $p < .05$ ** $p < .01$

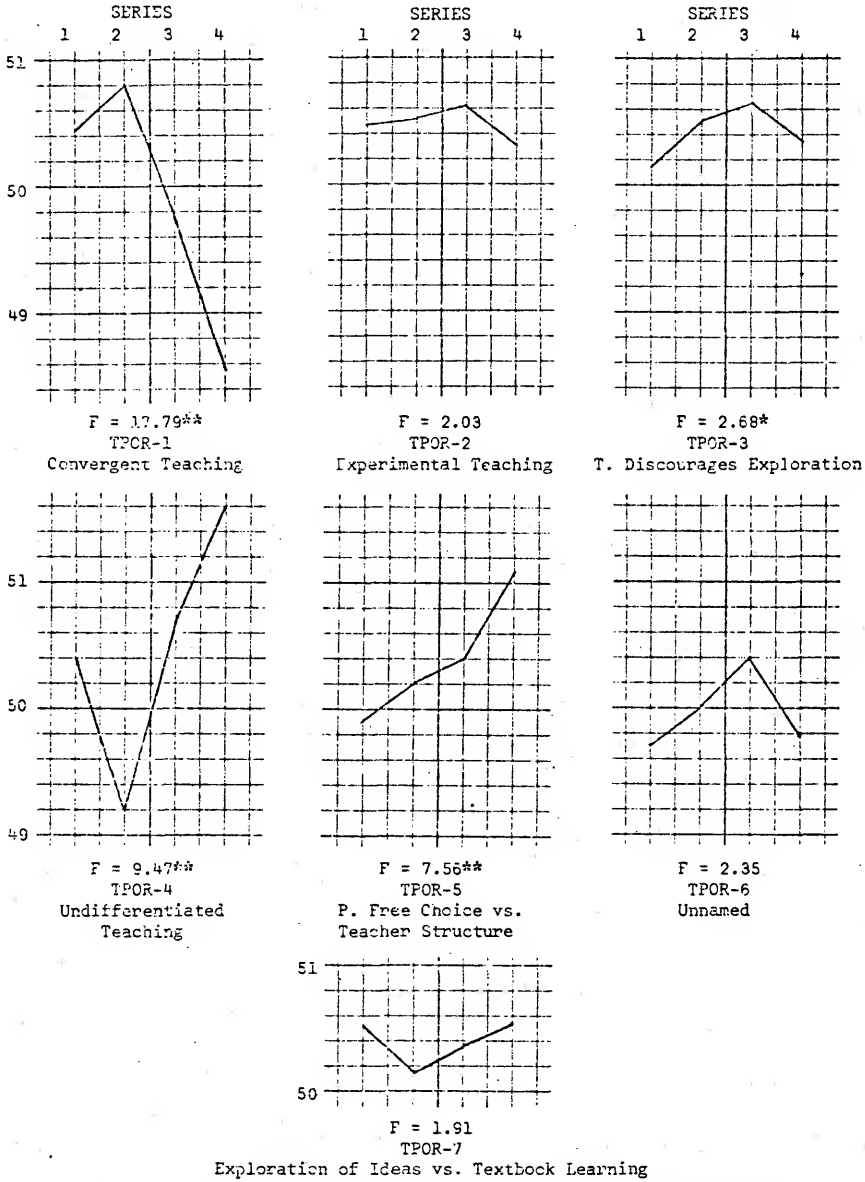

FIGURE 2. Graphs of mean factor scores for 289 classrooms across four series of observations during the day on each of seven TPOR factors.

Teaching, and Factor 5 - Pupil Free Choice vs. Teacher Structured Activity. Most convergent teaching occurs in the early part of the day falling off rapidly toward the end of the day. The most differentiated teaching occurs in the early part of the day with undifferentiated teaching high later in the day. Teacher structured activities tend to occur in the morning with increasing freedom of pupil choice over the course of the day.

Factor 3 - Teacher Discourages Exploration is also significant with regard to change over the day. The items of which this factor is composed are all odd-numbered items on the TPOR, i.e., they are behaviors attributed philosophically to the traditional teacher. All the items represent an attempt to keep the student "on the right track" in a teaching-learning situation. This behavior increased over most of the day and declined over the final series of observations.

Factor 2 - Experimental Teaching and to a large extent Factor 7 - Exploration of Ideas attempt to measure more experimental types of behavior, perhaps those behaviors which fit into the open-classroom, child-centered approaches to early education. Analysis of the measurement of such behaviors does not disclose significant changes over the time of day.

In general, the data from the TPOR agree with the data from the FLACCS in showing that most of the "work" (in the traditional sense) of the classroom tends to be done in the first half of the school day with students getting increased freedom of choice as the day goes on.

Observer Effects

The distribution of the sum of the ratings by two observers of the scale on classroom attitude is given for the 289 Follow Through Classrooms

in Table 7. The distribution for the scale on attention to observer is given in Table 8.

Incomplete factor scores for 289 classrooms, blocked (1) by the classroom attitude scores and (2) by the attention to observer scores, were subjected to two-way analyses of variance. Results are presented separately for each of the two scales.

TABLE 7

SUM OF RATINGS BY TWO OBSERVERS
ON THE SCALE OF CLASSROOM ATTITUDE

Group	Frequency	Sum of Two Ratings	Rating Description
			(1) Hostile
		3	
		4	(2) Uptight
1	13	5	
2	20	6	(3) Neutral
3	30	7	
4	70	8	(4) Polite
5	77	9	
6	79	10	(5) Friendly, Open

Classroom Attitude

Mean factor scores for the six groups of classrooms blocked by ratings on classroom attitude are given in Table 9 for the nine FLACCS factors and in Table 10 for the seven TPOR factors. Highly significant differences were found among classrooms with different ratings of

TABLE 8

SUM OF RATINGS BY TWO OBSERVERS
ON THE SCALE OF ATTENTION TO OBSERVER

Group	Frequency	Sum of Two Ratings	Rating Description
		2	(1) A Focus of Attention
1	10	3	
2	16	4	
3	34	5	
4	63	6	(3) A Normal Disruption
5	67	7	
6	62	8	
7	25	9	
8	12	10	(5) Ignored

classroom attitude on FLACCS factors 1, 5, 7, and 9. The up-tight to neutral end of the attitude scale was associated with more strong control (and teacher negative affect), less gentle control, more pupil negative affect and less teacher positive affect. The friendly end of the scale was, conversely, associated with less strong control, more gentle control, less pupil negative affect and more teacher positive affect.

Significant interactions between groups and time of day occurred for FLACCS factors 7 and 8 and TPOR factors 1 and 6. Mean factor scores for each group over the four periods of the day are given in Tables C-1 to C-4, respectively, in Appendix C.

TABLE 9

MEAN FACTOR SCORES ON NINE FLACCS FACTORS FOR SIX GROUPS OF CLASSROOMS BLOCKED BY RATINGS ON CLASSROOM ATTITUDE

FLACCS Factor	Group						F's		
	1	2	3	4	5	6	Groups	Time	Inter-Action
1	55.0	52.7	52.2	50.6	49.3	49.3	11.04**	15.10**	.91
2	51.0	49.4	49.3	49.5	50.5	51.0	1.60	7.36**	.82
3	48.8	49.6	49.9	49.3	50.3	51.2	1.52	2.80*	.53
4	48.7	48.6	50.4	50.4	50.1	50.1	1.07	1.19	1.29
5	47.8	49.8	49.6	49.2	50.3	51.6	3.23**	5.72**	1.08
6	50.6	51.3	49.7	49.8	50.1	50.5	.64	4.38**	.89
7	53.4	51.5	50.6	50.2	49.5	50.5	4.54**	16.04**	2.62**
8	49.4	49.8	49.2	50.8	49.7	50.2	.79	.72	1.76*
9	45.7	48.4	47.4	47.8	50.3	52.4	10.25**	.22	.63

* $p < .05$ ** $p < .01$

TABLE 10

MEAN FACTOR SCORES ON SEVEN TPOR FACTORS FOR SIX GROUPS OF CLASSROOMS BLOCKED BY RATINGS ON CLASSROOM ATTITUDE

TPOR Factor	Group						F's		
	1	2	3	4	5	6	Groups	Time	Inter-Action
1	48.4	49.8	50.5	51.0	49.3	49.6	1.38	13.77**	1.75*
2	49.9	50.9	49.7	49.4	50.8	51.5	2.41*	.84	1.22
3	51.0	50.9	51.4	51.5	49.8	49.6	1.94	.58	1.31
4	50.6	52.8	51.8	49.9	50.4	49.6	1.91	3.23*	.85
5	51.5	49.8	49.5	49.5	50.8	51.1	1.65	5.99**	.90
6	49.7	50.5	49.4	49.9	49.6	50.5	.88	.87	1.96*
7	50.2	50.2	49.2	49.2	50.9	51.3	1.87	.72	.86

* $p < .05$ ** $p < .01$

Attention to Observers

Of the TPOR factors, only TPOR-2, Experimental Teaching, was significantly related to grouping or classroom attitude. Classrooms that were rated most open and friendly were those in which the most experimental teaching occurred.

A similar procedure was followed for blocking classrooms on the basis of attention to observers as was followed for the ratings of classroom attitude. Results from a two-way analysis of variance for the nine FLACCS factors are given in Table 11 and for the seven TPOR factors in Table 12.

Again, significant differences were found among classrooms with different ratings of attention to observers on FLACCS factors 1, 5, 7, and 9. Greater amounts of attention to observers tend to be associated with more negative affect and strong control on the part of the teacher (FLACCS-1) and with more negative affect on the part of the pupils (FLACCS-7). Little attention to observers tends to be related to gentle control (FLACCS-5) and teacher positive affect (FLACCS-9). These results parallel the findings from classroom attitude ratings although they are neither as consistent nor as striking.

More of the behaviors of TPOR-6, the unnamed factor, occur in the extremes of the attention scale than occur in the middle ranges. The difference is significant but uninterpretable. There are four significant interactions of groups by time of day in TPOR factors 1, 3, 4, and 5. Mean factor scores for each of the eight groups over four periods of the day are given for TPOR factors 1, 3, 4, and 5 in Tables C-5, C-6, and C-8, respectively, in Appendix C.

TABLE 11

MEAN FACTOR SCORES ON NINE FLACCS FACTORS FOR EIGHT GROUPS OF CLASSROOMS BLOCKED BY RATINGS OF ATTENTION TO OBSERVERS

FLACCS Factor	Group								F's		
	1	2	3	4	5	6	7	8	Groups	Time	Interaction
1	52.5	52.8	51.9	50.5	50.2	48.8	51.1	49.8	3.02**	13.59**	.68
2	50.9	50.7	50.9	50.0	50.1	50.2	49.8	49.5	.51	5.41**	.47
3	50.2	50.4	50.3	49.6	50.1	50.0	51.1	51.2	.50	2.57	1.03
4	49.2	50.4	50.6	50.4	50.1	49.2	50.5	49.4	.45	1.96	.61
5	50.1	49.1	49.9	49.3	49.9	51.3	50.5	53.0	2.34*	3.04*	.96
6	52.2	50.4	49.5	49.9	49.9	50.4	52.1	48.9	1.98	3.89**	.70
7	53.1	51.7	51.9	50.0	50.1	49.1	51.2	50.3	2.94**	10.73**	1.21
8	48.5	50.6	50.1	50.7	49.7	50.2	48.8	50.0	1.04	4.54**	.89
9	48.8	47.4	48.6	48.8	49.7	50.8	50.9	51.3	2.10*	.76	.79

* $p < .05$ ** $p < .01$

TABLE 12

MEAN FACTOR SCORES ON SEVEN TPOR FACTORS FOR EIGHT GROUPS OF CLASSROOMS BLOCKED BY RATINGS ON ATTENTION TO OBSERVERS

TPOR Factor	Group								F's		
	1	2	3	4	5	6	7	8	Groups	Time	Inter-action
1	51.2	49.7	50.1	49.5	49.1	50.5	50.3	50.8	.54	20.58**	2.05**
2	51.1	50.7	50.8	50.1	50.4	50.0	52.3	50.6	1.35	.85	.95
3	49.8	51.5	50.2	50.6	50.4	50.4	50.3	50.0	.55	1.22	1.59*
4	49.6	51.2	50.5	50.6	50.7	49.9	50.5	51.1	.32	3.31*	1.66*
5	51.2	51.0	50.8	50.1	50.2	50.2	50.6	50.9	.27	10.23**	1.83*
6	51.1	50.6	49.6	49.7	49.3	50.0	50.9	52.2	2.83**	2.78*	.63
7	51.5	51.1	51.0	50.0	50.1	50.0	50.9	50.3	.56	3.24*	1.54

* $p < .05$ ** $p < .01$

CHAPTER V

DISCUSSION

Results have been reported for three related substudies investigating (1) the reliability of the FLACCS and the TPOR, (2) the stability of classroom behaviors measured by the two instruments, and (3) the effect of the observers in the classroom. Each of the substudies will be discussed separately.

Reliability

Intra-Observer and Inter-Observer Reliabilities

In the reliability data reported in the last chapter for nine FLACCS factors and seven TPOR factors, inter-observer or between-observer reliabilities are often noticeably poorer than intra-observer or within-observer reliabilities. In all but three cases, adjusted intra-observer reliability figures are higher than the actual overall reliability, while adjusted inter-observer reliability figures are lower. Even in the three cases where this does not hold true -- FLACCS-2, FLACCS-6, and TPOR-1 -- it is inter-observer reliability which has decreased the actual overall reliability figure, rather than intra-observer reliability. In any series of four observations, two each by two observers, only two combinations of series are possible for establishing intra-observer reliabilities -- the two series done by one observer and two series done by the other. Six combinations are possible, however, with four series

of observations. Only one combination -- series two and series three -- was selected to produce a measure of inter-observer reliability in this study. Three possible additional combinations were not calculated. Where the average of the expected reliabilities does not agree with the overall reliability, the discrepancy is due to inter-observer reliability figures which are not reported.

Comparison with Other Studies

Inter-observer reliability figures reported here, although in part a measure of agreement between observers, are not comparable to those measures of observer agreement in which different observers are tallying the same classroom behaviors at the same time. The inter-observer reliabilities reported here are based on different observers collecting data at two different times in each of 289 classrooms. Therefore, comparisons cannot be made meaningfully between the reliabilities reported in the present study and, for instance, the percent agreement scores based on a total instrument as reported by Stallings (1972), where two observers were paired in eight classrooms to collect data simultaneously.

Intra-observer reliabilities likewise have no equivalent counterpart in reported studies. Intra-observer reliabilities as reported by Brown (1968), for instance, are based on observations made by the same observers of behaviors from the same videotape a year apart. Intra-observer reliability as reported in the present study is based on observations by the same observer of the same classroom at two different times. Actual differences in behavior are reflected in the figures. Meaningful comparisons across the two studies cannot be made.

Implications from the Reliability Figures

Certain recommendations may be made in an attempt to increase reliability. Since inter-observer reliability figures were lower than intra-observer figures, additional training time might be spent observing those behaviors for which the inter-observer reliabilities were lower. An increase in reliability might also be gained by increasing the number of observations, although the corresponding increase in cost must also be considered. Changes in the observation instruments, themselves, might also help to increase reliability.

From the data presented it is obvious that certain groups of behaviors are recorded with more overall reliability than others. Behaviors related to teacher negative affect and strong control (FLACCS-1), pupil negative effect (FLACCS-7), teacher positive affect (FLACCS-9), or free versus more structured (FLACCS-2, TPOR-5, and TPOR-7) teaching styles are recorded with more reliability than those behaviors related to teacher gentle control (FLACCS-5), teacher attention in a task setting (FLACCS-8), undifferentiated teaching (TPOR-4) or the unnamed factor (TPOR-6). A closer look at these factors reveals an interesting finding. Those factors with reliabilities in the .80's are composed of more items of behavior than those factors with reliabilities in the .60's. Tables 13 and 14 show that higher reliabilities of the FLACCS and TPOR tend to be associated with larger numbers of items. If validity data (a significant correlation with pupil gain, for instance) indicated the importance of a factor and an attempt to increase reliability was deemed worthwhile, one method of increasing reliability could be to add to the observation instrument items which might be related to the factor. Adding items -- without increasing the total number of observations -- might increase reliability without an appreciable increase in cost.

TABLE 13

OVERALL RELIABILITY AND THE NUMBER
OF ITEMS FOR EACH OF NINE FLACCS FACTORS

FLACCS Factor	Overall Reliability	Number of Items
1	.84	23
9	.84	10
7	.83	13
2	.81	14
3	.78	11
4	.74	4
6	.70	3
5	.69	6
8	.64	5

TABLE 14

OVERALL RELIABILITY AND THE NUMBER
OF ITEMS FOR EACH OF SEVEN TPOR FACTORS

TPOR Factor	Overall Reliability	Number of Items
7	.82	16
5	.80	20
1	.78	7
3	.75	7
2	.71	13
6	.68	4
4	.62	3

The differences in reliability figures among FLACCS and TPOR factors demonstrate again the importance of calculating reliability figures for the variables used in research rather than calculating one reliability figure for the instrument as a whole. This point has already been made by Soar (1973) and by Johnson and Bolstad (1973).

Perhaps a further point needs to be made. Reliabilities based on data gathered in Follow Through classrooms may perhaps be higher than reliabilities using the same instruments in less variable classroom settings. The wide variety of approaches in the planned-variation phase of Follow Through increases the opportunity for variance across classrooms to be significantly greater than variance within. The definition of reliability as used in this study, is based on the concept of the average difference between two measures obtained in the same classroom being smaller than the average difference between two measures obtained in different classrooms. The possibility of higher reliability figures is increased as differences across classrooms increase. It is difficult to imagine a broader variety of kindergarten, first- and second-grade classrooms than exists in the present sample.

Stability of Classroom Behaviors

That much teacher behavior changes significantly throughout the school day is, perhaps, not surprising. Most kindergarten, first- and second-grade teachers have schedules or lesson plans which partition the day into periods of different activities. Indeed, what may be surprising is the similarity of the reported means across time of day for the factors of the FLACCS and the TPOR.

Although there were statistically significant differences in means across time for many classroom behaviors, those differences do not appear to be large. Significance was reached because of relatively small differences which were consistently replicated in other classrooms. Thus, significance was reached because much of the within-classroom variance could be accounted for by the time of day.

Despite significant changes in behavior over time, those behaviors may be measured reliably. The most significant changes in behavior in the present study occurred for teacher negative affect and strong control (FLACCS-1), pupil negative affect (FLACCS-7) and convergent teaching (TPOR-1). Overall reliabilities for these factors were .84, .83, and .78, indicating that instability of classroom behaviors during the school day need not necessarily be a barrier to good reliability. Lack of stability of classroom behavior does suggest, however, the importance of the time of day in an observation schedule. Predictable or lawful changes in some behaviors, e.g., structured teaching activities or negative affective behaviors, seem to occur at least at the kindergarten, first and second grade levels of the schools in the present study.

Observer Effects

Results of two-way analyses of variance -- time of day by groups blocked on classroom attitude or attention to observers -- were reported in Chapter 4. Those analyses were done to determine whether perceived differences in classroom attitude or attention to observers were accompanied by any significant differences in classroom behaviors. The most consistent of the significant differences among groups of classrooms blocked by ratings on classroom attitude and attention to observers

occurred in factors 1, 5, 7, and 9 of the FLACCS. The most significant associations involved FLACCS-1 and FLACCS-9 and the ratings on classroom attitude. The behaviors of FLACCS-1 are instances of teacher negative affect and strong control while behaviors of FLACCS-9 are indicators of teacher positive affect. The association of positive or negative teacher affective measures and ratings of classroom attitude are not surprising even though significant. Much of the information on which a rating of classroom attitude toward observers is made, is based on the teacher's affective behavior.

Samph (1968), in a study reported earlier, found that with an observer present in the classroom, the amount of criticism on the part of the teacher diminished significantly and the amount of praise increased. If, as Masling and Stern (1969) hypothesized, the effect of the presence of an observer diminishes gradually over time, then one might expect that the amount of criticism would increase and the amount of praise decrease over the course of a day of observation as teachers and students adjusted to the presence of the observer. From the present study, it is obvious that teacher negative affect (of which teacher criticism is a part) did, in fact, increase over the day. Positive affect (of which teacher praise is a part) did not decrease, however. Perhaps the presence of an observer differentially affects positive and negative affective behaviors. Perhaps changes in the amount of negative affective behavior for teachers and students over time are the result of something other than the presence of observers -- fatigue and irritability perhaps. Certainly, differences in the variables under consideration and differences in their measurement have complicated comparisons across the studies.

Of the TPOR factors, only TPOR-2, Experimental Teaching, was significantly related to the ratings on classroom attitude. Perhaps teachers who operate in open, experimental classrooms can more easily be friendly or open in their attitudes toward observers.

Implications for Further Research

Research using systematic observation would be strengthened by the reporting of reliability figures as demonstrated in this study. They emphasize, again, the wisdom of calculating reliability not on the total instrument, but on those variables used in the analysis of the data. Reliability figures point out the areas of relative deficiency within the observation instrument. They allow one to determine which areas need special emphasis during the training phase of a study. They also direct attention to areas where instrument revision might help to bolster reliability. They give the best overall indication of the accuracy of the measurements used in the data analysis phase of a study. Routine reporting of reliabilities computed in the data analysis phase of studies using systematic observation would allow greater opportunity for comparison of instruments as well as selection and refinement of instruments suitable for continued use.

Inter-observer reliability figures have implications for research design as well. Where between-observer differences are known to occur, individual observers should not be confounded with different experimental programs or treaments but should be distributed across all programs.

Results of the substudy on the stability of classroom behaviors have implications for the timing of classroom observations within the overall research design.

And, finally, further study is needed to investigate observer effects on classroom behaviors. The issue of reactivity has not been adequately researched.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Observable behaviors are the focus of systematic observation instruments. However, the data produced to measure classroom behaviors through systematic observation are the result of at least two categories of behavior: (1) that classroom behavior which is being observed, and (2) the behavior of an observer which results in tallies on the observation instrument. In addition, when more than one observer is involved in data collection, one must be concerned about (3) the behavior of additional observers. When more than one classroom is involved in the data collection, one is primarily concerned about (4) those behaviors which differentiate one classroom from another.

Questions about consistency arise with regard to each of the four categories of behavior: (1) Is a teacher consistent in her behavior? Is the class consistent in its behavior? (2) Is an observer consistent in recording what he sees? (3) Are all observers equivalent? To what extent are they the same measuring instrument? (4) Is the observation instrument capable of revealing systematic differences between classrooms and teachers?

In the present study, answers to the questions raised above were sought (1) by analyzing the stability of classroom behaviors, (2) by computing intra-observer (or within-observer) reliabilities, (3) by estimating inter-observer (or between-observer) reliabilities, and (4) by computing overall reliabilities on each of the measures of behavior.

The data under consideration were collected using two systematic observation instruments: the Florida Climate and Control Systems, FLACCS; and the Teacher Practices Observation Record, TPOR. Data from four series of observations, representing four periods throughout the school day, were collected in 289 Follow Through and comparison classrooms. Item data were T-scored by area transformation and reduced to incomplete factor scores on nine FLACCS factors and seven TPOR factors. From the data of a one-way analysis of variance on each of the FLACCS and TPOR factors, reliabilities were calculated. Results showed: (1) that many teacher behaviors change significantly over the course of a day while others remain quite stable, (2) that intra-observer reliabilities were, in general, high, (3) that inter-observer reliabilities were not as high as intra-observer reliabilities and that the inter-observer reliabilities of some factors were considerably better than others, and, (4) that overall reliabilities for each of the factors were generally in the .70's and .80's. Four of the 16 factors had overall reliabilities in the .60's. Significant changes in classroom behavior over the course of a day did not preclude high overall reliability in the measurement of those behaviors.

Results of several two-way analyses of variance demonstrated that some perceived differences in classroom attitude toward observers or attention to observers are accompanied by a significant differences in behaviors measured by FLACCS or TPOR factors. The most significant relationships were between teacher positive or negative affective behaviors and ratings of classroom attitude. More data are needed to properly assess the effect of observers on classroom behavior.

Implications for further research were drawn from the data of the study. Routine calculation and reporting of reliability data was recommended in all studies using systematic observation.

BIBLIOGRAPHY

Barr, A. S. Characteristic Differences in the Teaching Performance of Good and Poor Teachers of the Social Studies. Bloomington, Ill.: Public School Publishing Co., 1929.

Bellack, A. A., Kliebard, H. M., Hyman, R. T., and Smith, Jr., F. L. The Language of the Classroom. New York: Teachers College Press, Columbia University, 1966.

Brown, B. B. The Experimental Mind in Education. New York: Harper and Row, 1968.

Brown, B. B., Mendenhall, W., and Beaver, R. "The Reliability of Observations of Teachers' Classroom Behavior." Journal of Experimental Education, 1968, 36, 1-10.

Cronbach, L. J. Essentials of Psychological Testing. (2nd ed.) New York: Harper and Brothers, 1960.

Dopyera, J., and Lay, M. "Assessing the Program Environments of Head Start and Other Pre-school Children: A Survey of Procedures." Addendum to final report to OEO, Head Start Evaluation & Research, August, 1969. (Contract OEO 4120) Syracuse University (mimeograph).

Flanders, N. A. Teacher Influence, Pupil Attitudes, and Achievement. (Coop. Res. Monograph No. 12, OE-25040) Washington, D. C.: U. S. Department of HEW, 1965.

Flanders, N. A., "The Problems of Observer Training and Reliability." In E. Amidon and J. Hough (Eds.), Interaction Analysis: Theory, Research, and Application. Addison-Wesley, 1967.

Flanders, N. A. Analyzing Teaching Behavior. Reading, Mass.: Addison-Wesley, 1970.

Gage, N. L. "Research and Cognitive Aspects of Teaching." In The Way Teaching Is. Washington, D. C.: Association for Supervision and Curriculum Development and National Education Association, 1967.

Gordon, I. J. and Jester, R. E. "Techniques of Observing Teaching in Early Childhood." In R. Travers (Ed.) Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973.

Guertin, W. H., and Bailey, Jr., J. P. Introduction to Modern Factor Analysis. Ann Arbor, Mich: Edwards Brothers, Inc., 1970.

Hoyt, C. J. "Relations of Certain Correlation to Variance Ratio Estimates of Test Reliability." Twelfth Yearbook of the National Council of Measurements Used in Education, 1955, 50-55.

Hughes, M. Development of the Means for the Assessment of the Quality of Teaching in Elementary Schools. Salt Lake City: University of Utah Press, 1959.

Jackson, P. W. "The Way Teaching Is." In The Way Teaching Is. Washington, D. C.: Association for Supervision and Curriculum Development and National Education Association, 1967.

Jersild, A. T., Thorndike, R. L., Goldman, B., and Loftus, J. J. "An Evaluation of Aspects of the Activity Program in the New York City Public Elementary Schools." Journal of Experimental Education, 1939, 8, 166-207.

Johnson, S. M. and Bolstad, I. D. "Methodological Issues in Naturalistic Observation: Some Problems and Solutions for Field Research." In L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Eds.), Behavior Change: Methodology Concepts and Practice. Champaign, Ill.: Research Press, 1973.

Katz, L. G., Peters, D. L., and Stein, N. S. "Observing Behavior in Kindergarten and Pre-school Classes." Childhood Education, 1968, 44, 400-405.

Maccoby, E. E., and Zellner, M. Experiments in Primary Education: Aspects of Project Follow Through. New York: Harcourt-Brace-Jovanovich, Inc., 1970.

McGaw, B., Wardrop, J. L., and Bunda, M. B. "Classroom Observation Schemes: Where are the Errors?" American Educational Research Journal, 1972, 9, 13-27.

Masling, J., and Stern, G. "Effect of the Observer in the Classroom." Journal of Educational Psychology, 1969, 60, 351-354.

Medley, D. M. "Measurement Rationale." In Educational Testing Service, Disadvantaged Children and Their First School Experiences: Theoretical Considerations and Measurement Strategies. Princeton, N. J.: ETS 1968. PR-68-4 (2 vols.) Prepared under Contract OEO 4206 and Grant CG-8526, Office of Economic Opportunity.

Medley, D. M., and Mitzel, H. E. "Application of Analysis of Variance to the Estimation of the Reliability of Observations of Teachers' Classroom Behavior." Journal of Experimental Education, 1958, 27, 23-25. (a)

Medley, D. M., and Mitzel, H. E. "A Technique for Measuring Classroom Behavior." Journal of Educational Psychology, 1958, 49, 86-92. (b)

Medley, D. M., and Mitzel, H. E. "Measuring Classroom Behavior by Systematic Observation." In N. L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Mitchell, J. V., Jr. "Education's Challenge to Psychology: The Prediction of Behavior from Person-Environment Interactions." Review of Educational Research, 1969, 39, 695-721.

Mitzel, H. E. and Rabinowitz, W. "Assessing Social-Emotional Climate in the Classroom by Withall's Technique." Psychological Monographs, 1953, 67, (18, Whole No. 368).

Morsh, J. E. and Wilder, E. W. "Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900-1952." USAF Pers. Train. Res. Cent. Res. Bull., 1954, No. AFPTRC-TR-54-44.

Neujahr, J. L. "Classroom Observational Research." Educational Forum, January, 1972, 221-228.

Ober, R. L., Wood, S. E. and Roberts, A. "The Development of a Reciprocal Category System for Assessing Teacher-Student Classroom Verbal Interaction." Paper presented at the annual meeting of the American Educational Research Association, Chicago, February, 1968.

O'Leary, K. D. and Kent, R. "Behavior Modification for Social Action: Research Tactics and Problems." In L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Eds.), Behavior Change: Methodology Concepts and Practice. Champaign, Ill.: Research Press, 1973.

Ragosta, M. N., Soar, R. S., Stebbins, L. B. and Soar, R. M. "Sign Versus Category: Two Instruments for Observing Levels of Thinking." Paper presented at the meeting of the American Educational Research Association, New York, February, 1971. (Gainesville, Institute for Development of Human Resources, College of Education, University of Florida).

Rosenshine, B. "The Stability of Teacher Effects Upon Student Achievement." Review of Educational Research, 1970, 40, 647-662.

Rosenshine, B. Teaching Behaviors and Student Achievement. Windsor, Berkshire, England: National Foundation for Educational Research in England and Wales, 1971.

Rosenshine, B. and Furst, N. "The Use of Direct Observation to Study Teaching." In R. Travers (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973.

Samph, T. "Observer Effects on Teacher Behavior." Doctoral dissertation, University of Michigan, Ann Arbor, 1968.

Skindrud, K. "Field Evaluation of Observer Bias Under Overt and Covert Monitoring." In L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Eds.), Behavior Change: Methodology Concepts and Practice. Champaign, Ill.: Research Press, 1973.

Soar, R. S. An Integrative Approach to Classroom Learning. NIMH Project numbers 5-R11 MH 01096 to the University of South Carolina, and 7-R11 MH 02045 to Temple University, Philadelphia, 1966. ERIC document ED 033 749.

Soar, R. S. "Follow Through Classroom Process Measurement, Gainesville, Florida." Institute for Development of Human Resources, College of Education, University of Florida, 1971 Research report, Contract OEG-0-8-522471-4618 (100) to the University of Florida and OEG 0-8-522394-3991 (286) to the Florida Educational Research and Development Council, Gainesville, Florida.

Soar, R. S. "Final Report: Follow Through Classroom Process Measurement and Pupil Growth." Report to OE under Contract OEG-0-8-522394-3991 (286) College of Education, University of Florida, Gainesville, June, 1973.

Soar, R. S., Soar, R. M. and Ragosta, M. N. "Florida Climate and Control System: Training Manual." Institute for Development of Human Resources, College of Education, University of Florida, Gainesville, Florida, December, 1971.

Stallings, J. A. "Differences in the Classroom Experience of Follow Through and Non-Follow Through Children and the Relationship of Educational Processes to Child Outcomes on Test Scores." Symposium: Classroom Observation-Compensatory Education. Meeting at Stanford Research Institute, Stanford, California, September 7, 1972.

Thorndike, R. L. "Reliability." In E. F. Lindquist (Ed.), Educational Measurement. Washington, D. C.: American Council on Education, 1950.

Weick, K. E. "Systematic Observational Methods." In G. Lindzey and E. Aronson (Eds.), The Handbook of Social Psychology. (2nd ed.) Vol. II. Reading, Mass.: Addison-Wesley, 1968.

Westbury, I. "The Reliability of Measures of Classroom Behavior." Ontario Journal of Educational Research, 1967, 10, 125-138.

Withall, J. "Development of a Technique for the Measurement of Socio-emotional Climate in Classrooms." Journal of Experimental Education, 1949, 17, 347-361.

APPENDICES

APPENDIX A

TPOR AND FLACCS

TEACHER PRACTICES OBSERVATION RECORD

DIRECTIONS

The Teacher Practices Observation Record provides a framework for observing and recording the classroom practices of teacher. Your role as an observer is to watch and listen for signs of the 62 teacher practices listed and to record whether or not they were observed, WITHOUT MAKING JUDGMENTS AS TO THE RELATIVE IMPORTANCE OR RELEVANCE OF THOSE PRACTICES.

There are three (3) separate ten-minute observation and marking periods in each 30-minute visit to the teacher's classroom. These are indicated by the column headings I, II, and III. During period 1, spend the first five minutes observing the behavior of the teacher. In the last five minutes go down the list and place a check ( ) mark in Column I beside all practices you saw occur. Leave blank the space beside practices which did not occur or which did not seem to apply to this particular observation. Please consider every practice listed, mark it or leave it blank. A particular item is marked only once in a given column, no matter how many times that practice occurs within the ten-minute observation period. A practice which occurs a dozen times gets one check mark, the same as an item which occurs only once.

Repeat this process for the second ten-minute period, marking in Column II. Repeat again for the third ten-minute period, marking in Column III. Please add the total number of check marks recorded in Columns I, II, and III for each teacher practice and record in the column head TOT. There may be from 0 to 3 total check marks for each item.

_____
month day year

_____
Name of Teacher

_____
city state

_____
School

_____
Name of Observer-Judge

_____
Grade Subject

FIGURE A-1: THE TEACHER PRACTICES OBSERVATION RECORD

TEACHER PRACTICES OBSERVATION RECORD

TOT	I	II	III	TEACHER PRACTICES
				A. NATURE OF THE SITUATION
				1. T occupies center of attention
				2. T makes p center of attention.
				3. T makes some thing as a thing center of p's attention.
				4. T makes doing something center of p's attention.
				5. T has p spend time waiting, watching, listening.
				6. T has p participate actively.
				7. T remains aloof or detached from p's activities.
				8. T joins or participates in p's activities.
				9. T discourages or prevents p from expressing self freely.
				10. T encourages p to express self freely.
				B. NATURE OF THE PROBLEM
				11. T organizes learning around Q posed by T.
				12. T organizes learning around p's own problem or Q.
				13. T prevents situation which causes p doubt or perplexity.
				14. T involves p in uncertain or incomplete situation.
				15. T steers p away from "hard" Q or problem.
				16. T leads p to Q or problem which "stumps" him.
				17. T emphasizes idealized, reassuring, or "pretty" aspects of topic.
				18. T emphasizes realistic, disconcerting, or "ugly" aspects of topic.
				19. T asks Q that p can answer only if he studied the lesson.
				20. T asks Q that is not readily answerable by study of lesson.
				C. DEVELOPMENT OF IDEAS
				21. T accepts only one answer as being correct.
				22. T permits p to suggest additional or alternative answers.
				23. T expects p to come up with answer T has in mind.
				24. T asks p to judge comparative value of answers or suggestions.
				25. T expects p to "know" rather than to guess answer to Q.
				26. T encourages p to guess or hypothesize about the unknown or untested.
				27. T accepts only answers or suggestions closely related to topic.
				28. T entertains even "wild" or far-fetched suggestion of p.
				29. T lets p "get by" with opinionated or stereotyped answer.
				30. T asks p to support answer or opinion with evidence.

FIGURE A-1 (Continued...)

TOT	I	II	III	
				D. USE OF SUBJECT MATTER
				31. T collects and analyzes subject matter for p.
				32. T has p make his own collection and analysis of subject matter.
				33. T provides p with detailed facts and information.
				34. T has p find detailed facts and information on his own.
				35. T relies heavily on textbook as source of information.
				36. T makes a wide range of information material available.
				37. T accepts and uses inaccurate information.
				38. T helps p discover and correct factual errors and inaccuracies.
				39. T permits formation of misconceptions and over-generalizations.
				40. T questions misconceptions, faulty logic, unwarranted conclusions.
				E. EVALUATION
				41. T passes judgment on p's behavior or work.
				42. T withholds judgment on p's behavior or work.
				43. T stops p from going ahead with plan which T knows will fail.
				44. T encourages p to put his ideas to a test.
				45. T immediately reinforces p's answer as "right" or "wrong."
				46. T has p decide when C has been answered satisfactorily.
				47. T asks another p to give answer if one p fails to answer quickly.
				48. T asks p to evaluate his own work.
				49. T provides answer to p who seems confused or puzzled.
				50. T gives p time to sit and think, mull things over.
				F. DIFFERENTIATION
				51. T has all p working at same task at same time.
				52. T has different p working at different tasks.
				53. T holds all p responsible for certain material to be learned.
				54. T has p work independently on what concerns p.
				55. T evaluates work of all p by a set standard.
				56. T evaluates work of different p by different standards.
				G. MOTIVATION, CONTROL
				57. T motivates p with privileges, prizes, grades.
				58. T motivates p with intrinsic value of ideas or activity.
				59. T approaches subject matter in direct, business-like way.
				60. T approaches subject matter in indirect, informal way.
				61. T imposes external disciplinary control on p.
				62. T encourages self-discipline on part of p.

FIGURE A-1 (Continued...)

INSTITUTE FOR THE DEVELOPMENT OF HUMAN RESOURCES
University of Florida
Gainesville, Florida
November 25, 1970

FLACCS*

Florida Climate and Control System

Program_____ Teacher_____

City_____ Date_____

School_____ Observer_____

Grade_____ Series_____

*This is an experimental form which should not be cited or used without permission of the developers.

FIGURE A-2. THE FLORIDA CLIMATE AND CONTROL SYSTEM

NEGATIVE AFFECT

					Verbal	Teacher						Nonverbal
C.	Tot	1	2	3		C.	Tot	1	2	3		
10					Says "stop it," etc.	10						Waits for child
11					Uses threatening tone	11						Frowns
12					Rejects child	12						Points, shakes finger
13					Criticizes, blames	13						Pushes or pulls, holds
14					Warns	14						Shows disgust
15					Yells	15						Takes material
16					Scolds, humiliates	16						Refuses to respond to child
17					Other	17						Other
18					Code Involvement							

					Verbal	Pupil						Nonverbal
19					Says "No," "I won't," etc.	18						Makes face, frowns
20					Teases	19						Pouts, withdraws
21					Laughs	20						Uncooperative, resistant
22					Tattles	21						Stamps, throws, slams
23					Commands or demands	22						Interferes, threatens
24					Makes disparaging	23						Takes, damages property
25					Demands attention	24						Picks at child
26					Makes someone "feel small"	25						Pushes or pulls, holds
27					Finds fault	26						Hits, hurts
28					Threatens	27						Is left out
29					Other	28						Other
30					Code Involvement							

					Verbal	Teacher						Nonverbal
31					Says "Thank you," etc.	29						Accepts favors for self
32					Agrees with child	30						Waits for child
33					Supports child	31						Gives individual attention
34					Gives individual attention	32						Warm, congenial
35					Warm, congenial	33						Listens carefully to child
36					Praises child	34						Smiles, laughs, nods
37					Develops "we feeling"	35						Pats, hugs, etc.
38					Is enthusiastic	36						Sympathetic
39					Other	37						Other
40					Code Involvement							

					Verbal	Pupil						Nonverbal
41					Says "Thank you," etc.	38						Helpful, shares
42					Sounds friendly	39						Leans close to another
43					Agrees with another	40						Chooses another
44					Initiates contact	41						Smiles, laughs with another
45					Offers to share, cooperate	42						Pats, hugs another
46					Supports another	43						Agreeable, cooperative
47					Is enthusiastic	44						Enthusiastic
48					Praises another	45						Horseplay
49					Helps another	46						Other
50					Other	47						(Continued over)
51					Code Involvement							

CODE INVOLVEMENT
0. None involved
1. Few involved
2. Up to 1/2 the class
3. More than half

FIGURE A-2 (Continued...)

C.	Tot	1	2	3	TEACHER
10					Teacher Central
11					Leads singing,games,storytm
12					Moves freely among pupils
13					Withdraws from class
14					Uses blackboard, A-V Equip.
15					Ignores,refuses to attend P.
16					Attends P. briefly
17					Attends P. closely
18					Attends P. in succession
19					Attends simultaneous activ.
					VERBAL CONTROL
20					Praises
21					Asks for status
22					Suggests, guides
23					Feedback, cites reason
24					Questions for refltiv.thot
25					Correct w/o criticism (SM)
26					Questions for control
27					Questions states beh.rule
28					Directs with reason
29					Directs w/o reason
30					Uses time pressure
31					Call child by name (EWS)
32					Interrupts Pupil,cuts off
33					Warns
34					Supv. p. closely,imblizes.
35					Criticizes
36					Orders, commands
37					Scolds, punishes
38					Uses firm tone
39					Uses sharp tone
					NONVERBAL CONTROL
40					Tolerates deviant beh.
41					Positive redirection
42					Nods, smiles for control
43					Positive facial feedback
44					Uses "body English"
45					Gestures
46					Gives tangible reward
47					Touches, pats (gentle)
48					Holds,pushes,spanks(harsh)
49					Takes equipment, book
50					Signals, raps
51					Shhh! Shakes head
52					Glares, frowns

C.	Tot	1	2	3	PUPIL
10					Pupil Central
11					Pupil -- no choice
12					Pupil -- limited choice
13					Pupil -- free choice
14					(*Seat work w/o teacher
15					(*Seat work with teacher
16					(*Works, plays w. much supv.
17					(*Works, plays w. little supv.
18					(*Resists, disobeys directions
19					(*Obeys directions
20					Asks permission
21					Follows routine w/o reminder
22					Reports rule to another
23					Tattles
24					Gives information
25					Gives direction
26					Gives reason
27					Speaks aloud w/o permission
28					Engages in out-of-bounds beh.
29					Collaborates w. teacher
30					Task related movement
31					Aimless wandering
32					Fantasy
33					Uses play object as itself
34					Parallel play or work
35					Works,plays collaboratively
36					Works, plays competitively
37					Seeks reassurance, support
38					Shows pride
39					Shows fear,shame,humiliation
40					Shows apathy
					WORK GROUPS
41					Pupil as individual
42					Group w. teacher
43					Structured groups w/o t.
44					Free groups
					SOCIALIZATION
45					Almost never
46					Occasionally
47					Frequently
					MATERIALS
48					Structure T. behavior
49					Structure P. behavior
					PUPIL INTEREST ATTENTION
50-51					(Rank 1 low to 5 high

FIGURE A-2 (Continued...)

APPENDIX B

ITEM LOADINGS AND DESCRIPTIONS FOR FLACCS FACTORS 1 THROUGH 9 AND TPOR FACTORS 1 THROUGH 7

TABLE B-1

FLORIDA CLIMATE AND CONTROL SYSTEM

Factor 1 - Strong Control

Item	Loading	Description
11	.71	Teacher warns
12	.73	Teacher criticizes
13	.70	T orders, commands
18	.53	T holds, pushes, spanks (harsh)
19	.51	T says, "Shhh!" Shakes head
20	.77	T glares, frowns
26	.56	Pupil resists, disobeys directions
32	.46	P engages in out-of-bounds behavior
39	.41	P shows apathy
50	.62	Level 3, T verbal control
51	.82	Level 4, T verbal control
52	.87	Level 5, T verbal control
55	.50	Level 3, T nonverbal control
56	.70	T says "Stop it," etc.
57	.78	T uses threatening tone
58	.73	T criticizes, blames
69	.74	T frowns
70	.68	T points, shakes finger
71	.45	P makes face, frowns
72	.61	P uncooperative, resistant
82	.89	Total T negative
83	.52	Total P negative
1	-.42	Free groups

Eigenvalue = 11.69

*Source: Soar, 1973

TABLE B-2*

FLORIDA CLIMATE AND CONTROL SYSTEM

Factor 2 - Pupil Free Choice vs. No Choice

Item	Loading	Description
3	.41	T attends p closely
22	.62	Pupil limited choice
23	.60	Pupil free choice
33	.56	Task related movement
34	.63	P uses play object as itself
36	.59	P works, plays collaboratively
42	.67	Free work groups
44	.59	Frequent socialization
81	.43	P pats, hugs another
85	.50	Total p positive
2	-.65	Teacher central
10	-.54	T directs without reason
21	-.80	Pupil no choice
43	-.49	Almost no socialization

Eigenvalue = 6.95

TABLE B-3*

FLORIDA CLIMATE AND CONTROL SYSTEM

Factor 3 - Teacher-Pupil Supportive Behavior

Item	Loading	Description
5	.58	T suggests, guides
7	.43	T corrects without criticism (SM)
37	.59	P seeks reassurance, support
38	.57	P shows pride
48	.44	Level 1, T verbal control
62	.68	T supports child
67	.65	P agrees with another
77	.65	T waits for child (positive)
79	.55	P helpful, shares
80	.57	P leans close to another
85	.41	Total P positive

Eigenvalue = 5.73

*Source: Soar, 1973

TABLE B-4*

FLORIDA CLIMATE AND CONTROL SYSTEM

Factor 4 - Nonverbal Gentle Control

Item	Loading	Description
16	.64	T gestures
17	.47	T touches, pats, (gentle)
54	.65	Level 2, T nonverbal control
29	-.48	P gives information

Eigenvalue = 3.04

TABLE B-5*

FLORIDA CLIMATE AND CONTROL SYSTEM

Factor 5 - Gentle Control

Item	Loading	Description
6	.44	T questions for reflective thought
8	.65	T questions for control
14	.68	T nods, smiles for control
15	.54	T positive facial feedback
49	.65	Level 2, T verbal control
53	.70	Level 1, T nonverbal control

Eigenvalue = 3.44

*Source: Soar, 1973

TABLE B-6*

FLORIDA CLIMATE AND CONTROL SYSTEM

Factor 6 - Work Without Teacher

Item	Loading	Description
24	.58	P seatwork without teacher
35	.41	Parallel work or play
41	.62	Structured groups without teacher

Eigenvalue = 2.63

TABLE B-7*

FLORIDA CLIMATE AND CONTROL SYSTEM

Factor 7 - Pupil Negative Affect

Item	Loading	Description
30	.43	P gives directions
32	.63	P engages in out-of-bounds behavior
59	.70	P teases
60	.65	P commands or demands
61	.69	P threatens
71	.43	P makes face, frowns
72	.48	P uncooperative, resistant
73	.57	P interferes, threatens
74	.70	P takes, damages property
75	.66	P picks at child
76	.56	P pushes or pulls, holds
83	.75	Total P negative
43	-.41	Almost no socialization

Eigenvalue = 6.65

*Source: Soar, 1973

TABLE B-8*

FLORIDA CLIMATE AND CONTROL SYSTEM

Factor 8 - Teacher Attention in a Task Setting

Item	Loading	Description
3	.68	T attends P closely
25	.47	P works, plays with much supervision
40	.46	Pupil as individual (work groups)
46	.43	Materials Structure P behavior
47	.77	T closeness of attention

Eigenvalue = 3.13

TABLE B-9*

FLORIDA CLIMATE AND CONTROL SYSTEM

Factor 9 - Teacher Positive Affect

Item	Loading	Description
1	.52	Pupil interest attention rating
27	.49	P obeys directions
63	.46	T gives individual attention
64	.70	T warm, cogenial
65	.67	T is enthusiastic
66	.56	T sounds friendly
68	.62	P is enthusiastic (verbal)
78	.55	T smiles, laughs, nods
84	.78	Total T positive
85	.46	Total P positive

Eigenvalue = 5.23

*Source: Soar, 1973

TABLE B-10*

TEACHER PRACTICES OBSERVATION RECORD

Factor 1 - Convergent Teaching

Item	Loading	Description
1	.42	T occupies center of attention
21	.70	T accepts only one answer as being correct
23	.82	T expects p to come up with answer T has in mind
25	.70	T expects p to "know" rather than to guess answer to Q
27	.74	T accepts only answers or suggestions closely related to topic
45	.45	T immediately reinforces p's answer as "right" or "wrong"
47	.52	T asks another p to give answer if one p fails to answer quickly

Eigenvalue = 4.42

*Source: Soar, 1973

TABLE B-11*

TEACHER PRACTICES OBSERVATION RECORD

Factor 2 - Experimental Teaching

Item	Loading	Description
14	.47	T involves p in uncertain or incomplete situation
16	.67	T leads p to Q or problem which "stumps" him
18	.48	T emphasizes realistic, disconcerting, or "ugly" aspects of topic
24	.43	T asks p to judge comparative value of answers or suggestions
26	.46	T encourages p to guess or hypothesize about the unknown or untested
32	.51	T has p make his own collection and analysis of subject matter
34	.56	T has p find detailed facts and information on his own
38	.71	T helps p discover and correct factual errors and inaccuracies
40	.61	T questions misconceptions, faulty logic, unwarranted conclusions
42	.46	T withholds judgment on p's behavior or work
46	.42	T has p decide when Q has been answered satisfactorily
48	.52	T asks p to evaluate his own work
50	.65	T gives p time to sit and think, mull things over

Eigenvalue = 5.37

*Source: Soar, 1973

TABLE B-12*

TEACHER PRACTICES OBSERVATION RECORD

Factor 3 - Teacher Discourages Exploration

Item	Loading	Description
9	.52	T discourages or prevents p from expressing self freely
13	.66	T prevents situation which causes p doubt or perplexity
15	.65	T steers p away from "hard" Q or problem
35	.45	T relies heavily on textbook as source of information
43	.59	T stops p from going ahead with plan which T knows will fail
49	.52	T provides answer to p who seems confused or puzzled
55	.64	T evaluates work of all p by a set standard

Eigenvalue = 3.91

TABLE B-13*

TEACHER PRACTICES OBSERVATION RECORD

Factor 4 - Undifferentiated Teaching

Item	Loading	Description
51	.92	T has all p working at same task at same time
53	.57	T holds all p responsible for certain material to be learned
52	-.92	T has different p working at different tasks

Eigenvalue = 2.71

*Source: Soar, 1973

TABLE B-14*

TEACHER PRACTICES OBSERVATION RECORD

Factor 5 - Pupil Free Choice vs. Teacher Structured Activity

Item	Loading	Description
2	.61	T makes p center of attention
4	.74	T makes doing something center of p's attention
6	.65	T has p participate actively
8	.68	T joins or participates in p's activities
10	.42	T encourages p to express self freely
12	.60	T organizes learning around p's own problem or Q
32	.52	T has p make his own collection and analysis of subject matter
36	.47	T makes a wide range of informational material available
44	.51	T encourages p to put his ideas to a test
54	.59	T has p work independently on what concerns p
60	.68	T approaches subject matter in indirect, informal way
1	-.50	T occupies center of attention
3	-.68	T makes some thing as a thing center of p's attention
5	-.67	T has p spend time waiting, watching, listening
11	-.61	T organizes learning around Q posed by T
19	-.50	T asks Q that p can answer only if he studied the lesson
25	-.45	T expects p to "know" rather than to guess answer to Q
31	-.48	T collects and analyzes subject matter for p
33	-.47	T provides p with detailed facts and information
59	-.69	T approaches subject matter in direct, business-like way

Eigenvalue = 8.39

*Source: Soar, 1973

TABLE B-15*

TEACHER PRACTICES OBSERVATION RECORD

Factor 6 - Unnamed

Item	Loading	Description
20	.52	T asks Q that is not readily answerable by study of lesson
29.	.44	T lets p "get by" with opinionated or stereotyped answer
56	.60	T evaluates work of different p by different standards
61	.45	T imposes external disciplinary control on p

Eigenvalue = 2.18

*Source: Soar, 1973

TABLE B-16*

TEACHER PRACTICES OBSERVATION RECORD

Factor 7 - Exploration of Ideas vs. Textbook Learning

Item	Loading	Description
10	.60	T encourages p to express self freely
22	.71	T permits p to suggest additional or alternative answers
24	.43	T asks p to judge comparative value of answers or suggestions
26	.59	T encourages p to guess or hypothesize about the unknown or untested
28	.58	T entertains even "wild" or far-fetched suggestion of p
32	.43	T has p make his own collection and analysis of subject matter
36	.45	T makes a wide range of informational material available
44	.43	T encourages p to put his ideas to a test
58	.56	T motivates p with intrinsic value of ideas or activity
60	.50	T approaches subject matter in indirect, informal way
62	.47	T encourages self-discipline on part of p
35	.43	T relies heavily on textbook as a source of information
41	.42	T passes judgment on p's behavior and work
45	.41	T immediately reinforces p's answer as "right" or "wrong"
57	.43	T motivates p with privileges, prizes, grades
59	.46	T approaches subject matter in direct, business-like way

Eigenvalue = 6.02

*Source: Soar, 1973

APPENDIX C

DATA ON SIGNIFICANT INTERACTIONS BETWEEN TIME OF DAY AND GROUPS BASED ON RATINGS OF CLASSROOM ATTITUDE OR ATTENTION TO OBSERVERS

TABLE C-1

MEAN FACTOR SCORES SHOWING THE INTERACTION*
OF CLASSROOM ATTITUDE GROUPINGS AND TIME OF
DAY FOR FLACCS FACTOR 7, PUPIL NEGATIVE AFFECT

Groups	Time of Day 1	2	3	4	Row Mean
1	51.05	52.35	55.86	54.50	53.44
2	49.80	52.00	51.48	52.79	51.52
3	49.89	49.60	50.75	52.06	50.57
4	49.42	50.30	50.67	50.55	50.24
5	48.47	50.13	49.82	49.58	49.50
6	49.86	50.83	50.44	50.81	50.49
Column Mean	49.44	50.53	50.68	50.85	

*$F = 2.62$; $p < .01$

TABLE C-2

MEAN FACTOR SCORES SHOWING THE INTERACTION*
OF CLASSROOM ATTITUDE GROUPINGS AND TIME OF
DAY FOR FLACCS FACTOR 8, TEACHER ATTENTION IN A TASK SETTING

Groups	Time of Day 1	2	3	4	Row Mean
1	49.34	48.11	50.85	49.20	49.37
2	50.99	50.99	48.16	48.92	49.76
3	48.44	48.89	48.77	50.64	49.18
4	50.10	51.30	51.05	50.69	50.78
5	48.99	50.55	49.55	49.54	49.66
6	49.99	51.51	49.82	49.44	50.19
Column Mean	49.63	50.74	49.87	49.84	

*$F = 1.76$; $p < .05$

TABLE C-3

MEAN FACTOR SCORES SHOWING THE INTERACTION* OF CLASSROOM ATTITUDE GROUPINGS AND TIME OF DAY FOR TPOR FACTOR 1, CONVERGENT TEACHING

Groups	Time of Day 1	2	3	4	Row Mean
1	50.77	50.72	46.79	45.24	48.38
2	50.55	50.08	49.57	48.91	49.78
3	50.51	50.53	50.95	50.10	50.52
4	51.34	52.15	51.07	49.26	50.95
5	49.73	50.38	49.05	48.01	49.29
6	50.35	50.19	49.38	48.43	49.59
Column Mean	50.47	50.77	49.76	48.53	

*F = 1.75; $p < .05$

TABLE C-4

MEAN FACTOR SCORES SHOWING THE INTERACTION* OF CLASSROOM ATTITUDE GROUPINGS AND TIME OF DAY FOR TPOR FACTOR 6, UNNAMED

Groups	Time of Day 1	2	3	4	Row Mean
1	51.29	50.65	48.62	48.15	49.68
2	49.39	50.54	51.18	51.06	50.54
3	49.34	49.39	49.66	49.33	49.43
4	49.55	50.29	50.04	49.63	49.87
5	49.27	49.22	50.47	49.44	49.60
6	50.32	50.54	50.88	50.34	50.52
Column Mean	49.73	50.01	50.36	49.77	

*F = 1.96; $p < .05$

TABLE C-5

MEAN FACTOR SCORES SHOWING THE INTERACTION*
OF ATTENTION TO OBSERVER GROUPINGS AND TIME
OF DAY FOR TPOR FACTOR 1, CONVERGENT TEACHING

Groups	Time of Day 1	2	3	4	Row Mean
1	54.21	54.17	49.66	46.63	51.17
2	50.58	50.91	48.71	48.70	49.73
3	51.01	50.95	50.16	48.48	50.15
4	50.32	50.19	49.00	48.62	49.53
5	49.45	50.64	48.87	47.57	49.13
6	50.66	50.41	50.81	49.96	50.46
7	49.94	50.99	51.70	48.54	50.29
8	52.19	52.56	49.65	48.73	50.78
Column Mean	50.47	50.77	49.76	48.58	

*$F = 2.05$; $p < .01$

TABLE C-6

MEAN FACTOR SCORES SHOWING THE INTERACTION*
OF ATTENTION TO OBSERVER GROUPINGS AND TIME
OF DAY FOR TPOR FACTOR 3, TEACHER DISCOURAGES EXPLORATION

Groups	Time of Day 1	2	3	4	Row Mean
1	51.79	50.91	49.03	47.51	49.81
2	51.18	52.15	51.39	51.41	51.53
3	49.96	49.86	50.56	50.24	50.15
4	50.37	50.65	50.38	50.96	50.59
5	50.11	50.44	51.07	50.14	50.44
6	50.01	49.95	51.23	50.60	50.45
7	49.44	51.27	50.42	49.88	50.25
8	49.90	51.01	48.93	50.06	49.98
Column Mean	50.18	50.52	50.69	50.38	

*$F = 1.59$; $p < .05$

TABLE C-7

MEAN FACTOR SCORES SHOWING THE INTERACTION*
OF ATTENTION TO OBSERVER GROUPINGS AND TIME
OF DAY FOR TPOR FACTOR 4, UNDIFFERENTIATED TEACHING

Groups	Time of Day 1	2	3	4	Row Mean
1	53.60	49.30	47.94	47.57	49.60
2	51.23	52.15	50.79	50.67	51.21
3	51.09	48.55	52.10	50.36	50.53
4	50.75	50.01	50.14	51.38	50.57
5	50.17	49.09	51.58	51.97	50.70
6	48.86	47.99	50.21	52.38	49.86
7	50.38	48.57	50.35	52.85	50.54
8	52.25	50.39	49.67	52.17	51.12
Column Mean	50.40	49.18	50.68	51.60	

*$F = 1.66$; $p < .05$

TABLE C-8

MEAN FACTOR SCORES SHOWING THE INTERACTION* OF ATTENTION TO OBSERVER GROUPINGS AND TIME OF DAY FOR TPOR FACTOR 5, PUPIL FREE CHOICE VERSUS TEACHER STRUCTURED ACTIVITY

Groups	Time of Day 1	2	3	4	Row Mean
1	49.72	49.09	52.01	53.89	51.18
2	50.74	50.74	50.73	51.88	51.02
3	49.78	50.59	50.58	52.43	50.84
4	49.45	49.87	50.57	50.50	50.10
5	49.59	49.85	50.23	50.97	50.16
6	49.97	50.67	50.38	49.89	50.23
7	51.03	50.92	49.27	51.36	50.65
8	50.08	48.87	51.52	53.01	50.87
Column Mean	49.88	50.19	50.44	51.08	

*$F = 1.83$; $p < .05$

BIOGRAPHICAL SKETCH

Marjorie Ragosta (nee Nystrom), a native of Boston, Massachusetts, is a graduate of Girls' Latin School of Boston. She has attended Boston University and the University of Florida, where she graduated with high honors in the College of Arts and Sciences in June, 1962 and earned her Master's degree in Psychology in 1963. She was elected to Phi Beta Kappa in 1962. During the 1962-63 academic year, she was a National Science Foundation Cooperative Graduate Fellow and an honorary Woodrow Wilson Fellow. After two years abroad, she became an instructor in Psychology through the Division of Continuing Education, University of Florida and returned to graduate school. From 1969-72 she was Project Manager for the National Evaluation of Follow Through through the Institute for Development of Human Resources at the College of Education, University of Florida. She is a research psychologist in the Educational Studies Division of Educational Testing Service, Princeton, New Jersey.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert S. Soar
Professor of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John M. Newell
Professor of Education and
Assistant Dean, Graduate School

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard J. Anderson
Professor of Psychology

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1974.

Dean, College of Education

Dean, Graduate School