DOCUMENT RESUME

ED 052 227 TM 000 626

AUTHOR	Garfunkel, Frank
TITLE	Productivity of Young Children: The Development of the Classroom Behavior Scales (CRB).
INSTITUTION	Boston Univ., Mass. Headstart Evaluation and Research Center.
SPONS AGENCY	Office of Economic Opportunity, Washington, D.C.
PUB DATE	Aug 70
NOTE	47p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Performance, *Behavioral Science Research, *Behavior Rating Scales, Classroom Environment, *Classroom Observation Techniques, Emotionally Disturbed Children, Factor Analysis, Grade 1, Individual Differences, Learning Disabilities, Performance Factors, Preschool Children, *Productivity, Retarded Children, Sex Differences, Student Behavior, Student Teacher Relationship, *Test Construction, Video Tape Recordings
IDENTIFIERS	*Classroom Behavior Scales, CRB

ABSTRACT

The rationale for and development of the Classroom Behavior Scales (CRB) are discussed in connection with applications on four separate samples of Head Start, kindergarten and first grade children. Studies of observer agreement, stability, structure, comparisons over sex, classes, conditions, observers and over time are presented. Results showed high observer agreement, non-significant repeated trail correlations, consistent and expected changes in scale means over time--Head Start to first grade, unstable factor structures over samples and over repeated observation of the same sample. Differences between classes were consistently significant. An application with an experimental program did not show significant differences, which was an expected result. Classroom observational strategies are compared to educational and psychological testing as ways of getting measures of children in schools. While the latter will, in general, put a premium on stability, as a condition of test reliability, the former will emphasize individual variability. Data from CRB applications is presented to support this contention. The key for the scales of the CRB is appended. (Author)

Special Education Department
School of Education
Boston University

PRODUCTIVITY OF YOUNG CHILDREN:

THE DEVELOPMENT OF THE CLASSROOM BEHAVIOR SCALES (CRB)

Research Report of the Boston University
Head Start Evaluation and Research Center

Frank Garfunkel
Principal Investigator

August 1970

Special Education Department
School of Education
Boston University

PRODUCTIVITY OF YOUNG CHILDREN:

THE DEVELOPMENT OF THE CLASSROOM BEHAVIOR SCALES (CRB)

Research Report of the Boston University
Head Start Evaluation and Research Center

Frank Garfunkel
Principal Investigator

Research reported herein was performed pursuant to a contract with the Office of Economic Opportunity, Executive Office of the President, Washington, D.C., 20506. The opinions herein expressed are those of the authors and should not be construed as representing the opinion or policy of any agency of the United States Government.

Contract No: OEO 4116

August, 1970

Table of Contents

Tables

ACKNOWLEDGEMENT

Field research that calls for daily work in classrooms over an extended period of time, depends on the strong support of teachers, assistants, community workers and parents. More often than not we got not only support, but active and necessary involvement. As is discussed in the body of the report, such support and involvement is not only a question of convenience, but of theoretical necessity if we are interested in change -- not simply description.

Although the responsibility for the views herein is clearly ours, this research would have been meaningless without the involvement of Head Start, Follow Through and regular elementary school teachers, supervisors, principals and trainees from Boston Head Start, particularly the South End, Cambridge Head Start and Follow Through, and Brockton, Massachusetts Head Start and elementary schools. We gratefully acknowledge our indebtedness to all of the above school personnel, as well as to parents and children in these programs.

PRODUCTIVITY OF YOUNG CHILDREN:

THE DEVELOPMENT OF THE CLASSROOM BEHAVIOR SCALES (CRB)

Research Report of the Boston University
Head Start Evaluation and Research Center

Frank Garfunkel
Principal Investigator
Boston University

ABSTRACT

The rationale for and development of the Classroom Behavior Scales (CRB) are discussed in connection with applications on four separate samples of Head Start, kindergarten and first grade children. Studies of observer agreement, stability, structure,comparisons over sex, classes, conditions, observers and over time are presented. Results showed high observer agreement, non-significant repeated trial correlations, consistent and expected changes in scale means over time -- Head Start to first grade, unstable factor structures over samples and over repeated observation of the same sample. Differences between classes were consistently significant. An application with an experimental program did not show significant differences, which was an expected result.

Classroom observational strategies are compared to educational and psychological testing as ways of getting measures of children in schools. While the latter will, in general, put a premium on stability, as a condition of test reliability, the former will emphasize individual variability. Data from CRB applications is presented to support this contention. Furthermore, tests are developed with "success" in terms which mean successful individuals. This is clearly a form of institu-

tional bias which perpetuates existing conditions of potential and opportunity. Attention to classroom process will set up a more direct connection between intervention and effect than exists in traditional designs which tend to ignore or control for instructional variables. If goals derived from process oriented studies favor certain groups it will be more apparent than when those goals are obscured by the mysteries of test development.

Efforts in the direction of classroom behavior measurement call for different research methodologies, designs, measurement techniques, budgets and technology. Since they must necessarily be closely tied to teaching and learning in schools, they will have a more immediate effect on educational practices, and they will more immediately be effected by them. Therefore, process research will be involved in settings and will necessarily be directly associated with change.

I. Introduction

Although there have been many efforts to both describe and study the correlates of attitudes (Bowers, Davis, and Bowers, 1962; Travers and others, 1961), styles (Hughes, 1959), methods (Hayes, 1966), characteristics (Ryans, 1960; Heil and Washburne, 1962; Shim, 1965), and specific behaviors (Kounin and Gump, 1961) of teachers, there have been few efforts to focus on individual child behavior in classroom situations (Perkins, 1964, 1965). The latter calls for an extensive observational program over an extended period of time, as well as clearly stipulated dimensions which can apply to diverse situations, activities, and grade levels. Those investigators who have attended to classroom dynamics have generally limited themselves to highly specific behaviors such as language (Bellack and others, 1966) or kinds and amounts of social interactions (Amidon, 1966; Flanders, 1965). Because these variables are so specific to particular kinds of classes, ages of children, and subject matter, they are difficult to generalize across grade levels, subject matters, styles of classroom organization or school systems. They are even more difficult to apply as part of a training model.

In order to explore an attentive approach to observing, recording, and describing "educational" behavior of children, we have focused on productivity of children in classroom situations, which includes any and all parts of the school day, such as field trips and times when children are eating, resting, or in transition.

It is assumed that the impact of an educational experience on a child will be a function of the child's productivity in a succession of class days. Assessment of productivity involves judgments about what is taking place as well as the intensity of experiences for each child.

The measurement of intelligence involves linearly combining scores on a variety of items, each of which implies judgments about selection and criteria for success, but none of which is essential to a "high" score. The concept of intelligence, as operationalized in intelligence tests, is useful partly because it does not depend on highly specific abilities or reactions. Similarly, successful productivity can take place in a variety of ways and in many kinds of situations.

While productivity is here conceived of as a dependent variable -- dependent upon teacher, other children, classroom atmosphere, interactional patterns, and curriculum -- it is possible to raise questions about patterns of productivity in individual children across time, within classes, between teachers and between different school environments. Our preliminary work and our review of research in this field indicate that there is no simple connection between pupil performance and either teacher attitudes, or the behaviors of teachers in classroom situations, which have come under the headings of style, methodology, or instructional modes (Cogan, 1959; Harvey, 1967). The variability of child behaviors in different classroom situations and with different teachers is more or less accepted, but the implications of this have rarely affected research or practice (Cronbach and Snow, 1969). There is much talk about individual differences of both teachers and children, and there is glib acceptance of the axiom that particular classrooms might be better for some children than for others, but placement of children in classrooms is generally immutable. This research has specifically addressed itself to questions about variance and invariance of child behavior over time and in diverse educational settings.

II. Rationale

In any given time segment, a child's classroom behavior can be characterized as being more or less involved and at a higher or lower level of functioning. It is assumed that a classroom that generates higher levels of productivity for longer periods of time will have greater educational impact on its members. Although there is considerable evidence to the contrary, it remains to be seen whether methods, materials, or teaching style bear any direct relationship to productivity of individual children in any given sample of classes (Cogan, 1959; Harvey, 1967).

Quantification of productivity involves measurement of the extent of involvement in particular activities in which children are engaged. Ascertaining levels of productivity involves evaluation of the process in which children are engaged, and is not dependent upon curriculum -- specific materials being used, subject matter being studied or method being applied -- age of children, size of class, presence of teacher or other children, or whether or not it is a prescribed or proscribed activity.

Any child behavior can be looked at both in terms of form and content. We can ask questions about the structure of an interaction, a play activity, a lesson, or a lunch period. We can also inquire about the content of any of these activities. In order to have sufficient generality, a scale which is to be applied to such diverse phenomena must be sufficiently abstract so that it is not dependent upon specific characteristics of any one kind of form or content.

As a basic model for these scales, we have drawn heavily upon Guilford's "dimensions of intellect" (Guilford, 1957), particularly those that he calls "products" and "operations." Form that is relatively mechanical and which depends upon authoritative transactions is considered to be low level. Forms which are reciprocal and which depend upon experimentation, dialogue, and accommodation are considered to be of high level. Content that consists principally of labels, definitions, and memorization, whether it be colors and shapes or intellectual history, is of low level. On the other hand, content that involves transformations over time, space or people, implications or analyses is considered high level.

In both form and content there is a developmental progression from simple, unitary and one way processes to more differentiated, reciprocal and abstract processes. The application of the concept of productivity to educational process will require constant operational redefinition of dimensions of form and content in order to establish the unitary nature of the concept with regard to diverse educational situations.

In addition to ascertaining amount and quality of productivity that takes place in classroom situations, it is necessary to determine conditions under which it takes place, both in order to determine whether or not productivity is independent of other factors, and to give it a more valid meaning by placing it in its proper context. Conditions include : control -- who initiates activities and who is responsible for them being carried out; size of group -- including other children and teachers, direction and kinds of social interactions that are transpiring, and curriculum that is manifest at the time.

Thus, in any given time segment, individual children in classes are involved in processes which can be observed, and resulting recordings can be made which indicate levels of productivity. Furthermore, we can judge the extent of child involvement, continuity of the extent and kind of involvement over time, and the relative variation of children's productivity in any particular class, over time. We can observe and describe the conditions under which this productivity takes place in order to provide measurements of variables which will describe both individual child and class behavior in such a way as to bear a direct relationship to the assessment of educational impact. This descriptive process is not dependent upon teaching styles, methods, or materials, but can be systematically related to them in order to throw light on questions about appropriate placements for children in given educational situations.

Productivity is here conceived of as a dependent variable -- dependent upon the quality of contact that a teacher makes with a child individually or with a group of children, and the interaction of that contact with what individual children and teachers bring to classroom situations. It can also be conceived of as an independent variable which contributes to classroom

behaviors of children in ensuing classroom and out-of-school situations, and to cognitive, social, and emotional growth which can be measured either by the use of tests or through additional observational procedures.

Our goal is to be able to discuss atmospheres which are associated with greater and lesser productivity, but which are not amenable to any straight forward categorization, such as structured versus non-structured or democratic versus authoritarian. It may be that the only valid way to present data on atmosphere is through films.

There are several important assumptions behind this rationale which have affected the choice of an observational strategy and the content of that strategy. Foremost is the position that children can change their patterns of behavior in such a way as to drastically affect their ability to succeed in school and also to change standards of success. These changes cannot come about by the thoughtless application of an intellectually attractive methodology, the institution of a nation-wide program for poor children or the imposition of prescriptions which consider neither the teachers that use them or the children who are supposed to benefit from them. Changes in children will not be accomplished without careful attention to the educational process in which they are involved. The "little black box" approach, which includes a program which is presumed to operate in the same way on all children, accompanied by a series of pre and post tests, will inevitably lead us to the conclusion that children cannot change, although that is hardly the intended hypothesis that is being tested in most "black box" interventional studies. The real hypothesis that has been tested over and over again is whether any given interventional technique has a uniform effect on all children, which is, incidentally, confounded with the hypothesis that all teachers involved in an interventional program will administer it in the same way. The results of research on these two hypotheses are unequivocal (Barr, 1961; Heil and Washburne, 1962; Ryans, 1960; Shulman, 1970; Stephens, 1967; Turner, 1965). In addition to indicating that variations from interventional procedures are trivial compared to variation between teachers and individual children, they suggest that educational evaluation must concern itself with

variation within classrooms and within children rather than between programs, methodologies, and prescribed styles.

Therefore, in order to study and effect change, it will be necessary to address ourselves to those behaviors and situations that contain meaningful components of variability and which will provide a more direct and valid test of the hypothesis of change and of the potential for change.

A second critical assumption has to do with the relationship between data collected and agents who are directly involved with these data. We would question the collection of classroom data which are not or cannot be shared with teachers and which are designed and developed without their intimate and sustained involvement -- not because of a moral question but because of a scientific one. Teachers are links in the chain of change and are crucial to both its study and its accomplishment. To ignore them by creating instruments which are irrelevant to them or by designing research which simply uses them as objects of study will almost guarantee that they will not change and, consequently, that children in their classes will not change. Therefore, it is assumed that the rationale on procedures must be within the realm of understanding and application of teachers, not necessarily that they agree with it.

III. Procedures

Design

The variations and correlates of productivity have been examined across the following dimensions of independent variation:

grade level
experimental class versus regular class placement
observers
classes
time

The specific crossing of these dimensions has depended upon the characteristics of educational organization that was involved in each project.

The selecting of variables to be used as correlates of productivity has depended upon testing programs in the schools and agencies in which we worked, which included the national Head Start evaluation testing program.

Reliability of the procedure (observer agreement) has been systematically studied by having two observers rate the same child during the same time interval. The study of observer agreement has been built into each separate study because of changing observer personnel and the diversity of educational situations that have been used.

Instrument development has been an important part of the overall design. Variations in time segments, scheduling, observers, sampling of activities and variable definition have been used in order to further examine possibilities of the procedure.

Specific data on productivity has been obtained over the aforementioned conditions in order to provide a measure of the potential impact of an educational program on individual children and groups of children.

The procedures for assessing productivity have been applied to four samples of preschool (Head Start) and first grade children. Each application was aimed both at instrument development and at obtaining data which was to be used to more adequately describe variations in impact across children, teachers, and activities. In Brockton, Massachusetts, productivity of 56

Head Start children was assessed in seven classes where children had been randomly placed into classes, and teachers randomly assigned. Productivity recordings were again made on these same children during the spring and, again, during the fall of their first grade of public school in order to examine behavioral consistency of children in different classroom situations and over a period of a year.

A second application, using productivity as a principal dimension, involved an evaluation of eleven Follow-Through control and experimental classes in the Cambridge, Massachusetts, public schools (66 children.) This involved kindergarten and first grades in experimental and control classes with four observers and six observational periods.

Another use of the CRB was with Head Start classes in and around Boston, in order to examine variability of productivity and observer agreement.

A fourth application explained the correlates of productivity for 76 children in a testing program which included psychometric evaluation, parent interviews, sociometric evaluation and achievement tests.

This process of assessing productivity has been used on a limited basis in six classes for emotionally disturbed children, each of which was visited extensively by an experienced observer, in order to assess the validity of the instrument to get at classroom variations which had been assessed independently (Merrill, 196).

Result

These studies have shown that productivity, as will be operationally defined below, is a function of classroom environment; that there are consistent between-class differences in productivity; that observer agreement is such as to warrant the use of group data in evaluation studies (observer agreement correlations range between .60 and .92); and that productivity appears to be a unitary dimension which cuts across activities, ages, and teacher methodologies (or styles). Future applications of the process will be more specifically concerned with intellectually and emotionally disabled children who are in regular and special class settings, with particular emphasis on

preschool and primary grade level children, and evaluating productivity of children in classes which have been independently judged as being "high potential impact."

Instrumentation

Classroom Behavior Scales (CRB):

Scales have been developed to be used in classrooms as well as in classroom activities that take place outside of the physical environs of the classroom, in order to assess amount, kind, and conditions of productivity displayed by children of ages three to twelve (preschool and elementary.) This strategy focuses on child behavior in any and all situations. Ten-minute intervals are selected on a time sampling basis and records are made every thirty seconds. Fourteen scales are used to describe productivity of individual children. Scales and categories into which they fit are as follows:

1. amount of productivity
   a. participation: rejection to involvement
2. kind of productivity
   a. process focus -- form: authoritarian to experimental
   b. process focus -- content: mechanical to transformation
   c. use of materials: irrelevant to unique
   d. curricular activities
3. conditions of productivity
   a. control -- overt: teacher to child directed
   b. control -- covert: teacher to child directed
   c. behavior: withdrawn to hyperactive
   d. interaction: agents types and modes
   e. group size of specific activity being observed.

Observer agreement correlation (two observers recording on the same child at the same time) have varied between .60 and .92. Low correlations were found where there was restricted range in the use of the scales.

The CRB is a rating process rather than a rating scale. It involves an extensive training period for observers and it assumes a continuing flow of communication between observers and teachers. Ratings that are made are

routinely duplicated, given to teachers, and discussed so that a feedback process is always going on, depending upon the interest and willingness of the teachers involved.

Observers are required to have observed a class long enough so that they know the names of all children before any recording begins.

A series of ten-minute training films have been developed, each of which focused on an individual child in a setting similar to the one used. In addition, simulta-eous video taping and CRB recording have been done on a selected number of cases in order to provide a continuing check of observer reliability.

The key for the CRB, as well as the form used by observers, is included as an appendix.

Video Taping

Once we have determined who are the most productive children and classes, it becomes important to say something about the teaching which produces this productivity. We are dissatisfied with categories that have been used and are being used because they do not appear to hold up. Furthermore, we doubt that we can create useful categories at the present time because we have not arrived at a useful theoretical formulation which connects the various components of the teaching process. There are two possibilities for arriving at a more suitable independent variable which would be theoretically related to productivity and which would be useful to teachers, supervisors, trainers, and researchers, in gaining a more useful and greater understanding of the teaching process: (1) differentiation of teacher personalities (Jackson, 1963; Bjerstedt, 1967; Zimiles, Rabinowitz, and Hoy, 1964), (2) relation of variation in classroom atmosphere to patterns and amounts of child productivity (Sears, 1963). Both of these alternatives are beset by rather overwhelming measurement problems, as is well illustrated in the literature.

Given the difficulties of measurement, the strategy for developing a data base for relevant input variation must necessarily be fairly crude and must depend more upon the recipient than the giver. One approach to this is to develop carefully selected video tapes of classrooms that differ

markedly in the kinds and amounts of productivity that children display. These video tapes then would become basic data to be presented, along with data on productivity of individual children and of groups of children (Cogan, 1963). In addition, they can be used in order to train observers, provide a more complete feedback system for teachers and stimulate the formulation of hypotheses regarding the aforementioned two alternatives.

Sample

The numbers of children, classes, and disability areas used in these studies has depended upon negotiations with several public and private school systems. The goal has been to sample special and regular classes of preschool and elementary age children, including children who have been diagnosed as mentally retarded, emotionally disturbed, and learning disabled. Head Start classes have been sampled, including those in which disabled children have been placed and those in which they have not been placed..

IV. RESULTS, CONCLUSIONS AND SPECULATIONS

Observer Agreement

In connection with two applications, observer agreement was explicitly studied. Two observers completed ten-minute recordings for the same child, at the time. The team of observers was systematically rotated so that all observers were paired with all others. The first agreement study (1968) used six observers, making fifty matches (fifty children, a hundred protocols). Each CRB scale was analyzed separately. Correlations ranged from .23 to .92 (between first and second observers) with a median of .60. Scales with low correlations had restricted ranges of scoring responses -- on a five-point scale, only two or three points were used. The wording of these scales was revised in order to encourage use of the total range. In 1969, fifty matches were used with the revised scales. Correlations ranged between .60 and .92 with a median of .75.

Such levels of agreement were reached only after extensive training of observers, using video tapes and films, and careful study and delimitation of the scales. Although we were interested in idiosyncratic observer responses, we were also trying to develop a common language that would be useful for discussing the behavior of children in and out of classrooms. A rigid and excessive value put on observer agreement can eliminate perceptions which are critical to becoming aware of differences between children, classrooms and teachers. Disagreement between observers may mean that a given scale means different things to each, but it may also mean that they are seeing the situation in different ways and that both responses are precise. Further, it may be more important that the research reflects both views, than that it either eliminates one or compromises both in the name of observer agreement reliability. In subsequent research we will deal more directly with the relationship between styles of observation and behavioral styles of children and teachers. It may be that observer variations (rather than disagreement) will be a fertile source for the study of potentials for change. We would

expect that there is an important interaction between the observer and the behavior being observed which is, generally, either experimentally or statistically, controlled or cancelled out. It is often assumed, in the name of reliability, that observer agreement is synonymous with precision and that observer disagreement is either the result of ambiguous wording or poorly defined concepts. It is our contention that there is an important connection between observer disagreement and the potentials for change of teachers and children.

It is necessary here to distinguish between two types of disagreement:

(1) Disagreement is undesirable when discrepancies are attributed to misunderstanding of concepts and operations that are used in particular scales;

(2) Disagreement is desirable when there is basic understanding of the concept and operation of particular scales, but behavior in the classroom is seen in different ways because of important personality and value differences between observers.

An understanding of this second group of differences is critical to further instrument development where the principal focus is on how observational data can be used as a basis for feedback to teachers.

Although we did not get very far with this strategy in CRB studies, we did have one opportunity to apply the second kind of observer disagreement to a feedback model. We had two observers in the Brockton classrooms, both with small portable, battery-operated tape recorders. While they were in the classrooms, they continuously spoke in the recorders about what children were doing, what teachers were doing and what the general atmosphere of the classroom was like. Both of these observers had considerable experience teaching preschool children as well as children with intellectual and emotional disabilities, and both had specific training in observation. The tapes of their observations were transcribed. Each of these observers had his own purpose for recording observations, but a general project purpose was to provide teachers with feedback. Therefore, we encouraged teachers to read the transcripts of both observers who were in their classrooms. In this way, they were able to become directly and immediately aware of how different observers viewed what they were doing, the different language that was used

to describe similar behavior of the same children and teachers, and two different attitudes about the presence or absence of structure in their classrooms. This was an extraordinary demonstration of observer disagreement of the second kind, which teachers were able to accept. It resulted in extensive dialogue over a period of time with both observers, with parents and with their supervisors. We think that this kind of procedure should be formalized so that a narrow view of observer agreement synonymous with reliability will not sterilize observational data.

Stability

After a fair degree of observer agreement was established, we began asking questions about the stability of each scale and of the total CRB structure. How consistent are individual children and class groups over time? Is stability affected by time interval, classroom, observer or age? Does the multi-variate structure of the CRB remain constant over time, and for different samples?

Although it was not possible to set up designs to get unequivocal or even comfortable answers to these questions, we were able to deal with them in different ways. With the Cambridge Follow-Through sample (N = 66), three sets of observations were made on two separate days -- a total of six ten-minute recordings for each child. Means from each recording were correlated with means from other recordings for each scale. This resulted in 6 X 6 correlation matrices for each scale. The correlations for each scale within each day were uniformly positive and between .25 and .50. Correlations between days were uniformly non-significant (less than .25) with about equal numbers of positive and negative correlations. The participation, behavior and control scales were quite consistent -- they were also the scales that had the highest correlations of observer agreement. The process focus scales were somewhat more erratic. The interactional scales were extremely erratic and were extensively revised for later applications, as a result of these data.

The Brockton sample was followed for a full year, with CRB scales being used to collect data in the summer (during Head Start), fall (first grade) and spring (first grade) on the same group of children (N = 50). Six observations (two days) were made during the summer, nine (three days) in the fall and three (one day) in the spring. Observers had to achieve high standards of proficiency before they began formal observation.

Correlations for each scale, for summer x fall, fall x spring and summer x spring, were near zero (non-significant). CRB measurements did not reveal consistency of classroom behaviors of individual children between these three periods. They did reveal a fairly high degree of internal consistency for each of the scales.

The repeated testings of the Brockton sample over a one year period provided data on group stability. Mean scores of the sample in Head Start are consistently and significantly different from mean scores of the sample in first grade.

Table One

Means and Standard Deviations for CRB
Brockton Sample: Summer, Fall and Spring

Scale	Parameter	Summer N = 56	Fall N = 52	Spring N = 50
Control	Mean	3.51	2.60	2.35
	St. Deviation	.75	.27	.56
Process Focus: Form	Mean	1.79	2.27	2.34
	St. Deviation	.35	.34	.40
Process Focus: Content	Mean	1.73	3.49	3.57
	St. Deviation	.34	.30	.27
Participation	Mean	3.68	3.10	3.06
	St. Deviation	.39	.17	.23
Behavior	Mean	3.18	2.86	2.82
	St. Deviation	.30	.21	.26

It can be seen in Table One that control goes from the child, during Head Start, to the teacher during both CRB observations in the first grade; process focus, form and content both go up in the fall of the first grade and stay up in the spring of the first grade; participation gets less and behavior becomes more subdued during both first grade observations. The differences and directions are understandable in light of extensive anecdotally reported descriptions of Head Start and first grade classrooms in Brockton.

Looking at group differences in the Cambridge Follow-Through data, analyses of variance were done for each scale on each of the six trials, on the means and ranges of trials 1-3 and 4-6, and on the means and ranges of trials 1-6. The results of these analyses and more extensive description of results and interpretation can be found in the section on CRB variation below and in Table Five of that section. Classes were found to differ significantly and consistently on control, process focus-form, participation, style, group size and on most of the interaction variables. CRB data described classes uniquely, even if stability coefficients over individual children were near zero.

Structure

The definition of structure for an instrument that does not have demonstrable stability over individuals is a precarious undertaking, even if group data suggests stable group functions. Correlation matrices from each application show extraordinary and erratic variability, which is confirmed by factor analyses.. Factor structure is generally different for each sample, the only similarity being the common presence of what can be roughly termed as maturity and interaction dimensions. However, these can be interpreted as being artifacts of the two types of items which were in the CRB. Even in the respective Brockton matrices, calculated from data on the same children, structures are different for summer, fall and spring observations. A cursory inspection of correlation matrices shows some with many high r's, and others with relatively few. Average correlation coefficients for these

matrices range from .13 to .75 with a median of .23. The group of variables with the high estimated communalities vary considerably from one analysis to the other, both in those analyses using original samples (50 to 76 subjects) as well as in analyses of combined samples (n's = 152, 202 and 139).

Factor analyses were done on data from seven separate samples and three combinations of these samples. Three of the samples and one combination were recorded using the original CRB. The remaining samples and combinations were recorded using a revised form, as was discussed in the opening part of the previous section. The revised form included several reworded items -- in order to extend the range of responses -- and a different way of handling social interaction. Information on samples and combinations is presented in Table Two.

Table Two

CRB Samples for Factor Analyses

	Year	N	# of Variables	CRB Form	Average R	Sum H^2
Samples						
Cambridge Follow-Through (CFT)	68	66	17	Original	.18	8.60
Boston Head Start (BHS)	68	50	18	"	.12	7.53
Brockton Summer (BrS)	68	56	17	"	.24	10.68
Brockton Fall (BrF)	68	52	18	Revised	.35	11.93
Brockton Spring (BrSp)	69	50	18	"	.23	11.66
Boston South End (SE)	69	50	18	"	.54	14.89
Cambridge (Camb)	69	26	18	"	.75	16.28
Combined Samples						
BHS & CFT & BrS	68	152	18	Original	.13	7.68
BrF & BrSp	69	202	19	Revised	.19	11.20
SE & Camb	69	139	19	"	.20	11.04

Estimated communalities were inserted into the main diagonals of correlation matrices, all latent roots and factors were extracted, and those factors with roots greater than unity were rotated, using a normal varimax criterion. Rotations were done stepwise -- first, the two factors with the largest roots, then the three factors with the largest roots, etc.

For the three combined samples, the latent roots greater than unity and the percentages of trace accounted for are presented in Table Three; normalized communalities and three factor rotations for each scale are presented in Table Four.

Table Three

Latent Roots of CRB Factor Analyses

BHS & CFT & BrS N = 152		BrF & BrSp N = 202		SE & Camb N = 139	
Latent Root	% of Trace	Latent Root	% of Trace	Latent Root	% of Trace
3.03	40	3.84	34	4.18	38
1.58	21	2.67	24	2.53	23
1.21	16	2.06	18	2.15	20
	77	1.29	12		81
		1.00	9		
			97		

Table Four

Normalized Communalities and Three Factor Rotations
From Three Combined Samples

Scale	Samples: BHS & CFT & BrS				BrF & BrSp				Se & Camb			
		I	II	III		I	II	III		I	II	III
Quantity												
Participation	.73	x			.67		x		81	x		
Quality												
Process Focus Form	.71	x			.73				1.00	x		
Process Focus Content	.71	x			.73				1.00	x		
Use of Mat'ls	.35	x			.49				58	x		
Condition												
Control (Overt)	.49	x			.94	x		x	62	x	x	
Control (Covert)	na				.94	x		x	62	x	x	
Behavior	.73	x			.56		x		78	x		
Role of Child	na				.80	x			81	x		
Group												
Size	na				.46			x	.63		x	
Circle	na				.44				.51			
No. of 1's	na				.46				.72		x	x
Interaction Variables												
Style or Mode	.20				.91		x		67	x		
Style 0's	.37		x		na				na			
No Interaction	.57		x	x	na				na			
CH-ch Attention	.48				na				na			
CH-ch Verbal	1.00			x								
ch-CH Attention	.39											
ch-CH Verbal	1.00			x								
CH-T Attention	.46		x									
CH-T Verbal	.35											
T-CH Attention	.23		x									
T-CH Verbal	.57				na				na			
Agents	na				1.00	x		x	1.00			x
No. of 0's					1.00	x			.83		x	
No. of 1's					1.00	x			.83		x	
No. of 2's					.79			x	1.00			x
Type					.91	x	x		.52			
Appropriateness	na				.60				.57	x		

Notes: x indicates factor loadings greater than .4.
See Appendix for scale descriptions.
CH -- child being observed; ch -- other child; T -- teacher;
na -- not applicable.

In each of the three analyses, there is a factor that loads on process focus, participation, behavior and control; and two factors that load on interactional variables plus control. This speaks both to what has to be done about the instrument as well as to the internal structure of the CRB. There is not enough structural similarity accross samples to justify the use of factor loadings in scoring the instrument, or to make a direct connection between the CRB and inferences about the structure of classroom behavior.

Variations between samples confound this instrumental error with legitimate sampling variations. Included in the former is the observer training process and the selection of observers. We have not been able to carefully check out this confounding, but it has become clear that there is a discrepancy between scale and productivity variability. There was a dearth of classroom behavior variability in our samples, except for the relatively small South End and Cambridge Head Start samples. This meant that differences between observers would have an excessive effect. The more similar our samples of classes, the more differences will be due to observers, rather than reflections of real differences between classes. Similarly, the more heterogeneous the observers, the greater the differences between classes will have to be in order to come out significantly in an analysis. Obtaining samples of classes and observers which will allow between-class inferences to be drawn is a critical step in advancing our knowledge about educational variation.

These analyses show a consistent manifold of observed behaviors that includes cognition (process focus), motivation (participation), activity level (behavior) and independence (control). The usual bifurcation between cognition (aptitude-achievement) and social-emotional does not come out. Although the analysis of stability does not justify the use of CRB data as a measure of individual differences, the eventual possibility is suggested by the structure. However, we will probably have to begin conceptualizing such differences as individual-within-a-situation differences, which will vary both from individual to individual and from situation to situation, with an individual X situation interaction. One of the real problems of

these analyses is that we were stuck with our sample clusters. It is not at all clear that these sample clusters were comparable. The univariate comparisons of means and standard deviations presented in Table One can be easily interpreted, with the strong limitation that the interdependence of variables is not accounted for. We can say something about control and participation of the samples, but we cannot link them to one another, or to other dimensions of the space.

Multivariate comparisons require covariance regularity -- a given comparison can be meaningless because, disregarding central tendencies, there is restricted variance and, therefore, little or no covariance. The nature of classroom atmospheres is likely to lead to this kind of dilemna. The constellation of behaviors (CRB) in a classroom is generally either related or contingent. The more the teacher talks, the less the children talk. A high degree of observable participation will generally be associated with fairly high activity levels. This means that the proper description of classrooms is necessarily multivariate. But the set of contingencies might be quite different from classroom to classroom. We are used to stereotypes about teaching in classrooms -- which would make it appear that certain teacher characteristics might go together with certain kinds of classrooms. But our experience has been that there is no simple connection between the personalities and actions of teachers, and the classroom behavior of children. Although univariate connections obviously exist (Shea, 1968), there is a serious question about multivariate relationships. The factor analyses of our data cause us considerable concern in this direction. They suggest, because of slight structural consistency, fairly high observer agreement, but practically no individual child stability, that we have inadvertently chosen disparate samples of classes. The rules of the game are so different from class to class that the measurement format breaks down at the individual child level.

For example, we have to decide how many times and how long each time, to observe each child. So we begin observing many children in many classes and get data which suggests that the number of times a child has to be observed, depends on the child and the class, not upon a rule that will cover all applications of the instrument. But the number of times and

duration are important -- they tell us something else about the child. However, if we have to observe some children twice and others ten times, we are in a methodological dilemna which we do not know about until after the fact.

The point of this discussion -- or perhaps the apology -- is to articulate a basic dilemna of studying classroom behavior of children, as opposed to studying psychometric behavior of children or the classroom behaviors of teachers. The child functioning in the classroom is not strictly comparable to any other child functioning in any other classroom, in the same way that he is psychometrically comparable to other children. Classrooms and schools would appear to be different from one another in ways that affect child classroom functioning. The appropriateness of comparing child functioning will depend on the appropriateness of comparing classes. Until we can establish multivariate comparability of classes, measures of individual differences in classroom behavior will not mean anything and will produce the ambiguity that is evident in the factor analyses reported herein.

Putting this in a slightly different way, child classroom behaviors depend on the child and the classroom (teacher, other children, school). The first comparisons are those between children in the same class. Next, are comparisons between children in classes that have common structure, so that there is a meaning to looking for classroom behavior structure that is not interminably confounded with between-class variation. The next step will be to operationally define the dimensions of class structure. Lastly, individual differences in classroom behavior can be measured as a function of intervening structure.

CRB Variation over Sex, Conditions, Classrooms and Observers

Although we have many questions about the structural stability of the CRB, and its use as a measure of individual differences, several secondary analyses indicate that group comparisons are reasonable. This in no way validates the CRB, nor does it provide solid evidence for accepting or rejecting hypotheses about group differences. Rather, it provides an

empirical foundation for the future development of classroom behavior strategy, in general, and the CRB, in particular. We have already discussed CRB scale means and standard deviations of Brockton children when they were in Head Start and first grade (Table One). Changes in means are understandable in terms of observer reports. By using CRB data on the same sample over time, there is some modest element of control -- at a univariate level, we are beginning to tap relatively uncontaminated classroom behavior variation.

Because of the structure of the Cambridge Follow-Through program, it lent itself to secondary analyses of several independent variables. Because of arguments presented in the above section on 'structure,' univariate analyses of variance were used on each scale, for each trial and for three combinations of trials (1-3, 4-6, 1-6). For the latter, the dependent variables included means and ranges.

Analyses were done across the following dimensions:

sex
classes (11)
conditions (experimental X control)
observers (4)

Table Five presents obtained 'F' ratios for each CRB scale (original) across all dimensions except sex, which was uniformly non-significant. Analyses presented include only those done on means of trials 1-3, 4-6 and 1-6, which accurately represent the results of analyses of separate trials. The only exception to this was the analysis of variance for observers which is presented only for trials 1-3 and 4-6 -- the means for trials 1-6 for observers is not applicable because, in all cases, it includes two observers. The analyses of variance of ranges do not add anything and, therefore, are not presented.

Table Five

Cambridge Follow-Through: 'F' Ratios from CRB Analyses of Variance for each Scale across Classes, Conditions and Observers (N = 66) for Trials 1-3, 4-6 and 1-6

Dependent Variable				Independent Variable				
CRB Scales	Classes (11)			Condition (2) Exp X Control			Observers (4)	
Trials	1-3	4-6	1-6	1-3	4-6	1-6	1-3	4-6
Control	3.36	4.25	6.67	x	x	x	x	4.99
Process Focus Form	x	4.25	2.98	x	x	x	x	x
Process Focus Content	x	2.77	3.20	x	x	x	x	x
Participation	2.81	7.43	5.54	x	x	5.66	x	10.28
Behavior	2.66	x	4.04	4.79	x	4.91	x	x
Style	x	6.47	6.79	x	x	x	4.47	3.59
Use of Mat'l	x	2.74	4.28	x	x	4.75	x	x
Group Size Overall	x	9.19	10.90	5.08	27.53	21.27	x	x
Group Size Circled	2.24	3.80	7.09	x	10.96	10.72	x	4.70
Child-Child Interaction	2.34	3.66	5.69	x	x	x	5.90	x
Teacher-Child Interaction	x	2.71	2.22	x	x	x	x	3.59
Verbal Interaction	3.34	2.77	4.75	7.66	x	x	7.17	3.86
Attention	5.89	18.29	12.08	7.79	x	8.64	8.87	25.30
Subject Dominant	2.06	14.62	13.95	x	x	x	4.41	24.41
Other Dominant	8.57	3.06	3.03	5.29	x	x	23.96	6.53
Curricular Index	x	x	x	x	x	x	x	x
Degrees of Freedom		10,55			1,64		3,62	
Significance Level of F	p = .05	2.00			3.99		2.76	
	p = .01	2.66			7.04		4.13	

NOTE: x indicates F less than p = .05 level of significance.

It can be seen in Table Five that classroom 'F's' are consistently significant, particularly on the trial 1-6 comparisons. Our expectation for a scale of classroom behavior was that it would consistently and significantly distinguish classrooms of non-selected teachers. This expectation was fulfilled.

"Conditions" (Table Five) are generally not significantly different, with the exception of group size. As we reviewed that first year of Follow-Through, these results were to be expected. It was a new program which encountered great difficulties in getting started. In several schools, teachers would not willingly volunteer to be experimental teachers. The problem of developing a special program within the schools was disturbing both personnel and procedures. Our informal (CRB) and informal observations agreed -- Follow-Through did not make very much difference in classroom behaviors that first year. The differences in group size were straightforward effects of more adults and fewer children in experimental classrooms.

Differences across observers were most marked in the interaction scales -- but the observers had questioned the format of those particular scales and suggested revisions. These were made in a subsequent revision of the CRB which was used in later applications.

The pattern of results reported in Tables One and Five give us some support for our contention that the CRB is working at a group level. Such a strategy is basic to attending to what goes on in interventions -- the ways children actually function with a given teacher, methodology or class. Too much educational research has ignored classroom behavior and teacher effects while studying more accessible phenomena and characteristics. There has been much use of psychometrics, teacher questionnaires and sociometrics combined with the use of independent variables which were usually neither independent nor systematically variable.

Correlates of CRB Scales: Tests, School Grades

For the Brockton and South End-Cambridge samples, other data were available for children who had been observed and recorded with CRB protocols.

In the development of the CRB scales, there was no intention of validating them on psychometric or school success criteria. They were developed independently of these potential criteria for very definite reasons and biases -- educational and psychological tests and school grades are built on a structure of stability which is associated with reliability and social class. Measurement of school-related behaviors (as opposed to school behaviors) is an institutionalized way of keeping everyone on a steady course. Such measures have been validated by correlating them with the traditional success of successful people. When we think seriously about educational change, which is an imperative for lower income populations, we also have to ponder whether measurement procedures can either adapt to, or lead, the process. If schools are to be settings where change will be likely, we must be concerned with ways of altering existing behavioral patterns and of being more aware of the specific effects of alternative structures and personnel. Tests will certainly not fulfill this function. They can only be responsive to more efficient procedures of accomplishing traditional goals. But traditional goals and their accomplishment leave an important segment of school populations out of the running. Statistics on how many children are out of school and who they are, are very difficult to come by, but it is rather clear and accepted that many children drop out, particularly those from lower income minority groups who live either in inner cities, or in rural settings and who are often associated with migrant occupations. Without belaboring the point, the development of the CRB scales did not attend to such external criteria as tests and school grades. However, we thought it would be informative to look at correlations of CRB scales (or factor scores) with a number of tests and school grades.

For the Brockton sample, we had the following measures in addition to CRB's:

Lee Clark Reading Readiness Test
California Test of Mental Maturity
Peabody Picture Vocabulary Test
Average school grades during first grade

Correlations between these four measures and CRB scales were calculated for summer Head Start (1968), fall first grade (1968) and spring first grade (1969). Because all of the correlations were near zero, multiple correlations were calculated using psychometric measures and average grades as independent variables and each scale of the CRB as dependent variables. Multiple 'R's' were all very low and 'F' tests for residuals were non-significant with just one exception. CRB data was, therefore, independent of tests and school grades.

CRB factor scores were calculated for the Cambridge-South End data (N = 76) and correlated to the pre and post tests of the 1968-69 National Head Start Evaluation Program. The measures used in that evaluation were as follows:

Stanford-Binet Intelligence Scale
Animal House subtest of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI)
Preschool Inventory (Caldwell and Soule, 19)
Gumpgookies (Adkins and Ballif, 1967)
Sociometric Scale (Boger, 1969)

Factor scores were obtained from factors described in Table Four, above, with full knowledge of the tenuousness of the procedure. Correlations between factor scores and all of the above instruments were near zero (not significantly different from zero), which was the same result as that obtained from the Brockton data.

In spite of the fact that the development of the CRB did not consider aptitude and achievement tests, it was certainly surprising that no evidence for correlation was found in two separate samples, one a largely black sample from an inner city area, the other a white sample from a mill town about forty miles from Boston. The lack of stability for individuals may be all or part of the explanation for this. The cognitive measures, on the other hand, were fairly stable on test-retest -- Stanford-Binet Intelligence Scale, R = .68; WPPSI Animal House, R = .57; Preschool Inventory, R = .62. It is unlikely that the unstable CRB scales, with non-significant repeated measure correlations, would correlate significantly with other measures when they do not correlate with themselves.

Answers and Questions

Although our data provide no definitive evidence for the lack of relationship between classroom behavior and tests, it is worthy of some speculative discussion. Test theory and test construction have developed along well identified paths. They set up a behavioral situation, measure within it, and draw inferences for test results that ignore (or transcend) that situation. There is no question that test measures have been successful at going beyond test situations. From as little as a 15-minute group testing situation, a score can be obtained that will be highly predictive of how successful an individual will be in school -- which amounts to thousands of hours of behavior. While it is true that predictive power of testing is restricted to scholastic aptitude, certain industrial uses and a few clinical applications, their use is extensive. Whether testing contributes to stability (lack of change) or is simply a reflection of its existence in our society is a moot point -- in either case it should be quite clear that large-scale testing programs -- college entrance examination, graduate record examination, elementary and secondary school placement tests, military testing, civil service testing -- put a premium on behavior which is closely tied to criteria which were established on individuals who have already 'made it.' Equally important is the efficiency and format of the testing model -- take as little time as possible to make as big a prediction as is possible -- to make a prediction about a big thing that will take a long time. But this tells us something about who is to be the successful prototype -- someone who can show himself (maybe four years worth) in 15 minutes or two hours.

This study of the CRB shows that small samples of classroom behavior (30 minutes) are not predictors of other small samples of classroom behavior and, it follows, are not predictors of test behaviors. In other words, our evidence points out that measured classroom functioning is much more variable than measured test functioning. It can be argued that the consistency of tests is a function of their structure -- all children are asked the same questions in the same way -- and that the inconsistency of classroom behavior

is attributable to the measurement technique -- watch the child doing anything he happens to be doing. But this is precisely what the game of change is all about. If structures -- the what's and how's of questions or classrooms -- are held constant, there will be no change either over time or between conditions. Classroom behavior has characteristics which make it much more important, if not uniquely so, for change, than tests. One of these characteristics -- instability -- can be seen, at the same time, as a measurement headache, but also as a vital source of information about possible change, rather than about probable stability -- the usual standard error of a test score.

There are, then, a set of characteristics which are needed if we are to focus our attention on measurements which will facilitate change, rather than insuring a status quo. We have to be reasonably certain that quantitative and qualitative data can be communicated -- that operational definitions connect concepts and measurement to the mutual satisfaction of practitioners (in this case, teachers) and researchers. Legitimate instability cannot be sacrificed in the name of either reliability or convenience (money, time, effort, semester schedules, traditions of research funding.)

In the CRB we have developed an instrument which has passed some of the tests (sic) but which has a long way to go. We have continually involved teachers and parents in the development of the scales, including the choice of scales, verbalization, observational procedures and interpretation. A heterogeneous group of observers, including non-degree community people, teachers and doctoral students, have observed children together and communicated about behavior using CRB terminology. This agreement has been empirically verified by the observer agreement studies.

The CRB is not correlated with traditional test measurements, although this would appear partly a function of insufficient observational data. However, it is clear that we can get reliable (observer agreement) data from observing classroom behaviors which are not associated with test behavior. There are many ways in which the connections between classroom and test behaviors can be explored -- a search which is sorely needed. The fact that this connection has not been made, nor has it hardly even been looked into,

is an insidious comment on research design and strategies. The questions that are being asked, or so it would appear, are about what goes into the box and what comes out -- not what goes on inside. This clearly relegates process to insignificance. The process will continue to be ignored as long as psychometric canons prevail. Stick an I.Q. test with a dozen clinical reports and it will win out. Observational and clinical data will, in general, be obscured by tests and categories in any study which includes both, unless the tests and categories produce trivial results. The default of process means that process is given low order priority in the game of change -- for, if process were thought to be important to change, we would certainely be more interested in it. Or perhaps it should be stated the other way: if we were more resolutely interested in change, process would be our principal, and perhaps unique, concern.

We might begin asking questions about how we can change process, rather than how we can change individuals. This depends on whether we think changes in process can produce changes in individuals -- which does not mean that either process or individual transformations are mechanical, but that they take place in myriad ways, including mechanical. It was in terms of this line of inquiry that CRB scales were developed, applied and analyzed. But we have obviously only skimmed the surface of classroom process study. In other projects, we have filmed and video-taped teachers using a task structure -- each teacher does the same task so that many teachers can be compared in terms of process with task and materials held constant (Garfunkel, 1967b); participant observation methodology has been used to study how teachers deal with disturbance (Keane, 1969) and to compare experiences of black children who are bused to suburbs with those who stay in their inner city schools (Garfunkel, 1967a). The latter project used films as the principal media for presenting the final report (Garfunkel and Fiering, 1968) because the focus of that report was on process.

We see a need to simultaneously study process in a variety of ways, and to expend considerable energy on the development of vehicles for final reporting that will be appropriate to the subject matter and the audience. It is not at all clear that Fisherian statistics can do the job. Appropriate

mathematical models will have to be developed in order to deal with classroom realities, which are not at all like test realities. We certainely came nowhere near describing stable functions -- they are obviously more complex than univariate analyses will allow. It is difficult for generations of researchers steeped in analysis of variance, correlation, factor analysis and true scores to realize that there are grave discrepancies between our biologically derived experimental models and human learning in social situations. As we struggled with CRB data, the inappropriateness of many traditional analytical techniques became more and more obvious. There were times (more often than we would like to admit) when our analyses fitted the program library at our excellent computer center, which have been developed because of design and methodological traditions. But it became more and more obvious, as we got into our data, that it did not fit these traditions. Recent work by Cronbach and Snow (1969) takes a view of learning which is similar to ideas behind CRB development. As they found out, as is painfully obvious in this report, the schedules of time, effort and monies will have to be drastically revised in order to adequately describe instructional variables in close and continuing connection with individual pupil differences over time. But this also requires serious changes in school practices, which have developed as if instructional variables and their interaction with individual pupil differences are of negligible importance. We refer here to policies and practices relating to teacher training, hiring of teachers, supervision, testing of children, curriculum development, placement of children in classes and promotion policies.

Perhaps the most important innovation which would follow from attention to process would be the necessary concommittant opening up of systems. Derogation of process is perfectly consistent with the closed characteristic of most educational systems -- if process does not make too much difference, than, obviously, it does not have to be open for inspection, supervision or research. The book-keeping process cannot look at everything, so it keeps track of that which is most highly valued. Without drawing too many conclusions at the same time, it seems fairly evident that there are important

connections between values, programming and evaluation. Work on the CRB is an expression of values which would affect programming and how we view changing the educational order of perceptions.

APPENDIX

KEY FOR SCALES OF THE CLASSROOM BEHAVIOR FORM (CRB)

1 & 2 CONTROL: Overt & covert: Who controls the child's activities?

1. External: Teacher, other child or group; with direction and insistence.
2. External: Direction, suggestion without insistence.
3. Alternatives presented by teacher or other children.
4. External-Internal collaboration: Teacher-child or child-child discussion and participation in decisions.
5. Internal: Child controls his activities.

3. CURRICULUM:

Classification	Categories
1 Activity	11 Construction (wood, paper, clay, blocks, painting, collage, cooking, etc.)
	12 Performing (music, role playing, puppets, housekeeping, games, etc.)
	13 Play (water, sand, swing, jungle gym, jumping, etc.)
	14 Undefined (wandering, physically or otherwise)
	15 Interval
	16 Social Interaction Situations (casual play, laughing, talking, arguing, fighting, etc.)
2 Substantive	21 Science (biology, physics, chemistry, botany, zoology, etc.)
	22 Quantitative (numbers, sizes, shapes, puzzles)
	23 Language (speaking, discussion, stories, writing, reading)
	24 Social Relations (interpersonal relations, culture, social organization, history, community, etc.)
3 Routines	31 Snack (juice, lunch)
	32 Clean Up (of room, desks, tables, sink, etc.)
	33 Rest
	34 Other (arrival, departure, toileting, washing, dressing, etc.)

4. PROCESS FOCUS - FORM: At what level is the form (structure) of the activity? How does it take place?
   1. Traditional, authoritarian, force: Repetition, telling, direct questions, textual presentations, fighting.
   2. (Intermediary point)
   3. Demonstration, showing and illustration.
   4. (Intermediary point)
   5. Experimental, systematic procedures, planning, deduction, induction, explanations, with involvement and participation of child.

5. PROCESS FOCUS - CONTENT: At what level is the content? What is taking place?
   1. Mechanical: Labels, definitions, memorization, recitation.
   2. Skills: Motor, verbal.
   3. Precepts: Visual, auditory, tactual discriminations.
   4. Concepts: Ideas, issues.
   5. Transformations: Relationships between concepts, situations, events, implications, analyses.

6. PARTICIPATION: Involvement. How is the child participating in what he is doing?
   1. Rejection, Reluctance.
   2. Indifference.
   3. Mild Involvement.
   4. Moderate Involvement.
   5. Intense Involvement.

7. BEHAVIOR: Activity level. How active is the child?
   1. Very low, non-responsive, withdrawn.
   2. Low passive.
   3. Average.
   4. Moderately active.
   5. High: Acting out, aggressive.

8. INTERACTION. Agents.
   0. None
   1. Teacher and child(ren)
   2. Child and child(ren).

9. INTERACTION: Type (between child and other children or teacher and child.)
   0. None
   1. Parallel play or work.
   2. (Intermediary point)
   3. Non-verbal interaction.
   4. (Intermediary point)
   5. Verbal interaction.

10. INTERACTION: Mode of interactions.
    0. None
    1. Negative indication, fighting, restriction, disagreement, destructive.
    2. (Intermediary point)
    3. Neutral.
    4. (Intermediary point)
    5. Positive indication, supporting, sharing, reassuring.

11. USE OF MATERIALS: How teacher or child uses materials, equipment, games, facilities, people, situations.
    1. Irrelevant.
    2. Prosaic, repetitive.
    3. Conventional, routine.
    4. Appropriate, but with some diversity.
    5. Unique, resourceful, creative, unconventional.

12. ROLE OF CHILD:

    1. Spectator.
    2. (Intermediary point)
    3. Variable.
    4. (Intermediary point)
    5. Active participant.

13. GROUP SIZE: (include teacher) Number of individuals working or playing together. If group is clearly involved in common activity, circle number.

14. APPROPRIATENESS: Observer's subjective reaction to child's behavior or to interaction.
    1. Inappropriate.
    2. Neutral. (Does not have to be recorded.)
    3. Appropriate.

Bibliography

Adams, R. S. and B. J. Biddle. Realities of Teaching - Explorations with Video Tape. New York: Holt, Rinehart and Winston, 1970.

Amidon, E. Improving Teaching: The Analysis of Classroom Verbal Interaction. New York: Holt, Rinehart, and Winston, 1966.

Barr, A. S. Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness: A Summary of Investigations. Madison, Wisconsin: Dembar Publications, 1961.

Bellack, A. A. and others. The Language of the Classroom. New York: Teachers College Press, Columbia University, 1966.

Bjerstedt, A. Teacher Personality and Teacher Effectiveness. Malmo, Sweden: Department of Educational and Psychological Research, 1967.

Bowers, N. D., O. L. Davis and M. W. Bowers. The Effectiveness of the Index of Adjustment and Values in Predicting Classroom Behavior, Nineteenth Yearbook of the National Council on Measurement in Education, Ames, Iowa: National Council on Measurement in Education, 1962, pp. 112-120.

Cogan, M. L. The Behavior of Teachers and the Productive Behavior of their Pupils, Journal of Experimental Education, 1958, 27, pp. 89-124.

Cogan, M. L. Research on the Behavior of Teachers: A New Phase, Journal of Teacher Education, 14: pp. 238-243, Sept. 1963.

Cronbach, L. J. and R. E. Snow. Individual Differences in Learning Ability as a Function of Instructional Variables. U. S. Office of Education, Contract No. OEC 4-6-061269-1217. Stanford, California: Stanford University, March, 1969.

Flanders, N. A. Teacher Influence, Pupil Attitudes and Achievement. Washington: U. S. Department of Health, Education and Welfare, Office of Education, 1965.

Garfunkel,F. Observation of Teachers and Teaching: Strategies and Applications. Boston: Boston University, 1967a Distributed by Editorial Resources Information Center, ED-023-616.

Garfunkel, F. Teaching Style: The Development of Teaching Tasks. Boston: Boston University, 1967b. Distributed by Editorial Resources Information Center, ED-023-616.

Garfunkel, F. and A. Fiering. Reading is Fun: Black Kids in White Schools. Film developed by Boston University Head Start Evaluation and Research Center and Project Concern. Boston University, 1968.

Guilford, J. P. A Revised Structure of Intellect. Reports from the Psychological Laboratory of the University of Southern California, 1957, No. 19.

Harvey, O. J. and others. Teachers Beliefs, Classroom Atmosphere and Student Behavior. Final Report OEO-1274, ED 018-249, July, 1967.

Hayes, R. B. The Effects of Student Reactions to Teaching Methods. Harrisburg, Pennsylvania: Bureau of Research, Department of Public Instruction, 1966.

Heil, L. W. and C. Washburne. Brooklyn College Research in Teacher Effectiveness. Journal of Educational Research, 55: pp. 341-351, May, 1962.

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. Issue Editor. Teacher Personality. Theory Into Practice, 2: pp. 57-112, April, 1963.

Keane, A. A Study of Teaching Style in Relation to Children's Behaving. Boston University, Unpublished Master's thesis, 1969.

Kounin, J. S. and P. V. Gump. The Comparative Influence of Punitive and Non-punitive Teachers upon Childrens' Concepts of School Misconduct. Journal of Educational Psychology, 52: pp.44-49, February, 1961.

Merrill, M. Student-centered Teaching - An Approach to Teaching Emotionally Disturbed Children. Unpublished Master's thesis, Boston University, 1968.

Perkins, H. V. Classroom Behavior and Underachievement. American Educational Research Journal 2: pp. 1-12, January, 1965.

Perkins, H. V. A Procedure for Assessing the Classroom Behavior of Students and Teachers. American Educational Research Journal 1: pp. 249-260, 1964.

Ryans, D. G. Characteristics of Teachers. Washington, D. C.: American Council on Education, 1960.

Sears, P. S. The Effect of Classroom Conditions on the Strength of Achievement Motive and Work Output of Elementary School Children. U. S. Department of Health, Education and Welfare, Office of Education, Cooperative Research Project No. 873, Stanford California: Stanford University, 1963, 311 pp.

Shea, T. The Teaching-Learning Process in Observation Seminars. (Doctoral dissertation, Boston University). Ann Arbor, Michigan: University Microfilms, 1967

Shim, C. P. A Study of the Cumulative Effect of Four Teacher Characteristics on the Achievement of Elementary School Pupils. Journal of Educational Research, 1965, 59: pp. 33-34.

Shulman, L. S. Reconstruction of Educational Research. Review of Educational Research, 1970, 40-3: pp.371-396.

Travers, R. and others. Measured Needs of Teachers and Their Behavior in the Classroom. U. S. Department of Health, Education and Welfare, Office of Education, Cooperative Research Project No. 444. Salt Lake City: Department of Educational Psychology, University of Utah, 1961.

Turner, R. L. The Acquisition of Teaching Skills in Elementary School Settings: A Research Report. Bloomington, Bureau of Educational Studies and Testing, School of Education, Indiana University, 1965.

Zimiles, H., B. Biber, W. Rabinowitz, and L. Hoy. Personality Aspects of Teaching, A Predictive Study. Genetic Psychology Monographs 1964, 69: pp. 101-149.

SUPPLEMENTARY BIBLIOGRAPHY

Classroom Climate Studies

Gump, P. V. Classroom Behavior Setting - Its Nature and Relation to Student Behavior. Final Report, 1967, ED-015 515.

Kounin, J. S., P. V. Gump and J.J. Ryan III. Explorations in Classroom Management. Journal of Teacher Education 12: pp. 235-246, June, 1961.

Kowatrakul, S. Some Behaviors of Elementary School Children Related to Classroom Activities and Subject Areas. Journal of Educational Psychology, 50: Pp. 121-128, June, 1959.

Problem Solving Approach to the Problem of Teacher Effectiveness

Turner, R. L. The Acquisition of Teaching Skills in Elementary School Settings; A Research Report. Bloomington, Bureau of Educational Studies and Testing, School of Education, Indiana University, 1965.

Turner, R. L. Problem Solving Proficiency Among Elementary School Teachers. Bloomington, Institute of Educational Research, Indiana University, 1964. ED 001 276.

Turner, R. L. and Fattu, N. A. Skill in Teaching: A Reappraisal of the Concepts and Strategies in Teacher Effectiveness Research. Bulletin of the School of Education. (Indiana University) 36: pp.1-40, May, 1960

Role Approach to the Appraisal of Teacher Effectiveness

Biddle, B. J. and others. Studies in the Role of the Public School Teacher. Vol. I. Orientation, Methods and Materials. June, 1961. ED 001 408.

Rosencranz, H. A. and Biddle, B. J. The Role Approach to Teacher Competence. Columbia: Social Psychology Laboratory, University of Missouri, 1962, 42pp. (mimeo). (In Contemporary Research on Teacher Effectiveness, Biddle and Ellena, eds. New York: Holt, Rinehart and Winston, 1964).

Studies on Creativity and Productive Thinking

Gallagher, J. J. Analyses of Teacher Classroom Strategies Associated with Student Cognitive and Affective Performance. Final Report. 1968. ED-021 808.

SUPPLEMENTARY BIBLIOGRAPHY (Continued)

Spaulding, R. L. Achievement, Creativity and Self-Concept Correlates of Teacher-Pupil Transactions in Elementary School Classrooms. ED-024 463.

Spaulding, R. L. A Coping Analysis Schedule for Educational Settings. Durham, North Carolina: Educational Improvement Program, Duke University, 1966.

General

Shumsky, A. In Search of Teaching Style. New York: Appleton-Century-Crofts, 1968.