Portland State University

PDXScholar

Dissertations and Theses

Dissertations and Theses

7-28-1975

Development of an Observation Instrument to Assess Behavior

Scott Marshall Nay

Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the [Child Psychology Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Nay, Scott Marshall, "Development of an Observation Instrument to Assess Behavior" (1975).

Dissertations and Theses. Paper 2285.

<https://doi.org/10.15760/etd.2282>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

AN ABSTRACT OF THE THESIS OF Scott Marshall Nay for the
Master of Science in Psychology presented July 28, 1975.

Title: Development of an Observation Instrument to
Assess Behavior.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

Morris Weitman, Chairperson

Robert E. Jones, Jr.

Chad Karr

The Vernon-Nay Category Instrument was developed as the result of an experimental project in an inner-city school. The effects of the reorganization of the administration of the school and the total staff participation in a workshop about human relations and communications were evaluated by the change of behavior of fifth- and sixth grade students in the classroom. The Flanders' Interaction Analysis (assesses teacher-student verbal behavior) would measure change that occurred; however, the Vernon teachers wanted an assessment instrument that could, also, be used in

the future, without professional assistance. The F.I.A. observation system was too complex for the staff's, unassisted, use (statistical interpretation is based on the ratios of direct to indirect verbal teacher behavior, and student responses); consequently, many simpler (verbal and non-verbal) observation instruments were examined. No instrument was found to be appropriate for this particular situation. A new, non-verbal, instrument was designed that fit the needs of this staff and could also be used in adult-child situations other than the classroom.

Three times during the year, the V.N.C.I. and F.I.A. were used, concurrently, to observe and record the behavior of 180 students in 45 classrooms. Content validity of the V.N.C.I. was determined by teachers and other professionals. Nonparametric coefficients of correlation were calculated to estimate the (concurrent) validity of the V.N.C.I. by the results of the F.I.A. The inter-rater reliability of the V.N.C.I. was satisfactory, estimated by Scott's "pi" coefficient. Many nonparametric coefficients of correlation were calculated to assess change of behavior before and after the new programs were implemented; these coefficients were computed to correlate (grouped) categories of behavior, measured by both instruments, at each observation time of the year. While behavior did not change significantly (assessed by either instrument), the behavior desired by the staff was similarly assessed by the F.I.A. and V.N.C.I. at each time of the year.

The V.N.C.I. was re-examined, based on the observers' comments and the statistical results; several categories were changed or eliminated (V.N.C.I.-II). The V.N.C.I.-II positive categories were more highly correlated with the similar F.I.A. categories and inter-rater reliability of the V.N.C.I.-II was higher than with the V.N.C.I.

The staff believed the programs implemented were of value (irrespective of the instruments' assessments of the students' behavior) and have continued to use the V.N.C.I.-II, unassisted. The V.N.C.I.-II should be used in situations other than the classroom to further validate the new instrument. The F.I.A. makes the assumption that the verbal behavior of the teacher with the students is a random sample of either's behavior in the classroom. This assumption needs examination; the V.N.C.I. (I and II) non-verbal observations of the students correlated with the analysis of the F.I.A. for some verbal teacher behavior; however, the other V.N.C.I. (I and II) categories of student behavior did not result in significant statistical coefficients of correlation with the F.I.A. assessment of teacher or student verbal behavior.

DEVELOPMENT OF AN OBSERVATION INSTRUMENT
TO ASSESS BEHAVIOR

by

SCOTT MARSHALL NAY

A thesis submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE
in
PSYCHOLOGY

Portland State University
1975

TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the Committee approve the thesis of
Scott Marshall Nay presented July 28, 1975.

Morris Weitman, Chairperson

Robert E. Jones, Jr.

Chad Karr

APPROVED:

Ronald E. Smith, Head, Department of Psychology

David T. Clark, Dean of Graduate Studies and Research

August 1, 1975

ACKNOWLEDGMENTS

With gratitude, I express my appreciation to the staff of Vernon School. This staff convinced itself that new, untried solutions to very old problems were gambles worth taking. The decision to be creative and innovative, to risk the convenience of the status quo, was inspired by Walter Stickel. Without Mr. Stickel's "salesmanship" and his ability to be what he thought teachers should be, there would not have been an environment conducive to evaluating the effects of teachers' behavior by assessing change in students' behavior. I am especially grateful and proud of Dan McCall, Principal of Vernon School. Mr. McCall allowed decisions to be made and programs implemented that he was, ultimately, legally responsible for; his delegation of authority and commitment to a group decision-making model of administration are indicators of an extraordinary principal and person.

The expertise of my thesis committee was invaluable for this paper. They helped convert my planned evaluation into a research project worthy of submission as a thesis. Dr. Weitman's encouragement and ability to translate concepts into quality research brought this paper to fruition. I appreciate, too, the constructive critiques and practical solutions offered, with patience, by Dr. Karr and by Dr.

Jones, who served on my thesis committee.

I am very grateful for the efforts and abilities of my research staff, Sheri Teasdale and David Johnson; their personal observations were as important as their systematized observations. A special thank you is extended to my friend, Judy Bobair, who assisted with the completion of this paper.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS.....	iii
LIST OF TABLES.....	viii
 CHAPTER	
I INTRODUCTION.....	1
The Situation.....	3
II REVIEW OF EVALUATION LITERATURE.....	6
Organization and Staff Development...	6
National Training Laboratories Schmuck and Runkel Berenson Bigelow	
Non-Verbal Behavior Observations.....	12
Madsen Kowatrakul Craig and Holland Wandt and Ostereicher	
Conclusions.....	16
III DEVELOPMENT OF THE V.N.C.I.	18
Methods-Part I.....	18
Initial Field Tests.....	18
Categories Determined.....	20
Point-Time Sampling.....	21
Flanders' Interaction Analysis...	22

CHAPTER		PAGE
	Conclusions.....	25
	Methods-Part II.....	26
	Observer Training.....	26
	Subjects.....	28
	Procedure.....	30
	Expected Results.....	32
IV	STAFF IMPROVEMENT PROGRAMS.....	34
	Organizational Development (Shared Decision-Making).....	34
	Staff Development.....	40
V	RELIABILITY AND VALIDITY.....	45
	Reliability.....	45
	Inter-rater Reliability.....	45
	Flanders' Interaction Analysis..	47
	Vernon-Nay Category Instrument..	49
	Behavior Changes.....	54
	Validity.....	58
	Content Validity.....	58
	Concurrent Validity.....	58
VI	V.N.C.I.-II DEVELOPED.....	61
VII	CONCLUSIONS.....	65
	Research Results.....	65
	Future Research.....	67

PAGE

A SELECTED BIBLIOGRAPHY.....	69
LIST OF APPENDICES.....	71

LIST OF TABLES

TABLE		PAGE
I	Observation Schedule.....	29
II	F.I.A. Reliability Coefficients.....	48
III	V.N.C.I. Reliability	
	a) Verbal/Non-Verbal Reliability.....	50
	b) In/Out of Seat Reliability.....	51
	c) Twelve (Summated) Categories	
	Reliability.....	52
IV	I/D and S/I Ratios.....	55
V	I/D Ratios Ranked.....	56
VI	F.I.A./V.N.C.I. Rank Correlations.....	59
VII	V.N.C.I.-II Reliability.....	64

CHAPTER I

INTRODUCTION

I have worked in and around elementary school classrooms for six years and with groups of adolescents (locked up in jail) for two years. I have heard many group-workers and teachers (and occasionally myself) say, "This group is worse than last year's," and, "This group never pays any attention." When asked what these statements mean, the answer is usually that the group acts terrible or is more lethargic than some other group. When asked what has led to this conclusion, a frequent answer is, "I just know" or, "They had a better attitude." These are teachers' and group-workers' attitudes about children that other adults hear and accept as fact. Common talk in a lunch room is, "Wait until you get this group, they are horrible."

I believe adults use some form of behavior assessment when they label a group of children good or bad; often the teachers and group-workers I have discussed the subject with are not satisfied with their own answers, but they say they do not know how to gather relevant facts. An evaluation or research group is usually called to gather data for them. This can only occur when money is available.

My experiences with attitude scales, surveys, observation instruments, personality tests, and group dynamics has led me to the conclusion that adults need and want a simple, objective (and inexpensive) method to assess behavior.

As an employee of an elementary school, I was presented with a situation (and a small budget) that allowed me to explore the interaction of teachers, children and behavior changes. Five research projects were conducted by this school's staff and me. One project caused me to design and field-test an observation instrument that could be used in many group situations; situations where an adult works with a group of children and the adult wants factual information about the group's behavior.

I. THE SITUATION

Vernon Elementary School is a city school with eight grades and is located in Portland, Oregon. It is attended by approximately five hundred and fifty students. During the year of this study (1972-1973) it had a staff of sixty five; the staff consisted of twenty-five classroom teachers, ten specialists, twenty teachers' aides, and ten custodial or cafeteria personnel. The population of the school was: 50% Black, 40% White, and 10% other minorities. The socio-economic level of the community was, predominately, low income families; 67% of the students were classified "educationally disadvantaged" by federal guidelines.

Walter Stickel, a specialist in staff development and organizational problem-solving, was assigned to the school for one year as an administrative assistant. His general responsibility was to assist the principal in the administration of classroom/educational matters, and his specific assignment was to work with the staff to identify organizational problems and methods of eliminating these problems. He was also interested in evaluation tools and designs which could be employed to assess the total school program.

Stickel, working with many members of the staff, concluded that the general areas of concern were:

1. Staff feelings of inability to communicate with and understand the feelings of students;
2. ineffectual communication among the staff;
3. concerns about isolation from other members of the staff and an inability to share problems and solutions;
4. staff not having an established problem-solving mechanism;
5. staff not knowing what the specialists were doing and how best to utilize their help;
6. teacher inability to effect changes within the school or within the grades they were teaching;
7. knowledge of how students change their attitudes towards the school, the teacher, or themselves during the school year was based on personal opinions that could not be brought to consensus.

In response to these concerns, Stickel and other members of the staff developed long- and short-range procedures to deal with these problems. Stickel developed a plan for a staff reorganization and a staff communications workshop (Chapter IV). I did the research necessary to evaluate changes in the students' behavior during the school year relative to Stickel's programs. The staff agreed to provide funds for the research if the evaluation tools could be employed by the teachers, in later years, without professional assistance; \$1,000 was allocated for the research project.

My immediate task was to find previous studies relevant to the planned program. I assumed that this was not a unique approach to staff and organizational development and that an evaluation design existed that would be appropriate for this project.

CHAPTER II

REVIEW OF EVALUATION LITERATURE

I. ORGANIZATION AND STAFF DEVELOPMENT

National Training Laboratories. Organizational and staff development programs are an extension of the group development process first introduced by the National Training Laboratories in 1947. Argyris (1964) and Bennis (1962) have presented extensive discussions of these original programs. Organizational development programs are implemented to aid large organizations in accomplishing their mutual tasks. Many variations of the original programs, used extensively by Esso Company, have been developed.

Several studies have investigated the effects of organizational change in business and industry; only three studies were found that related to school reorganization.

Schmuck and Runkel (1971) reported results from an intensive study of a Beaverton, Oregon Junior High School. This school implemented an organizational development program based on the concept of organizational health discussed by Argyris (1964). The general outline of the program was to improve staff communication, task accomplishments, internal integration, and mutual adaptiveness of the organization to its environment. A detailed program was imple-

mented to accomplish these goals, but no evaluation procedure was used to measure the effects of the program. Reports from the staff indicated apparent classroom changes such as a better atmosphere in the classroom, or more attention to the feelings of the students. The different observations were noted but not measured systematically.

David Berenson (1964) conducted a study to explore the effects of a human relations workshop on the classroom performances of elementary school student-teachers. It was (Berenson, page 3, paraphrased) an integrated didactic and experiential human relations training program based on twenty-five hours of training in the discrimination and communication of the interpersonal conditions of accurate empathy, positive regard, genuineness, concreteness, immediacy, significant-other differences, and confrontation. The student-teachers were assessed before and after training on several rating scales. The effects on the classroom behavior of the students were not measured.

Berenson cites many other research projects with the same types of training and the effects of the training on the trainees. A recurring research design throughout his report, and that of Schmuck and Runkel's, was to initiate a training program or organizational program and then assess the effects on the trainees. The only research found that measured the effects of staff programs by change in students' classroom behavior was that of Bigelow.

Ronald Bigelow's (1970) research, "Effects of Organizational Development on Classroom Climate," had direct parallels in design to the proposed Vernon programs. Bigelow briefly discussed the history of organizational development and its possible effects in the classroom. With the achievement of the goals of the organizational development program in his research, the staff should be experiencing good organizational health. Organizational health, as defined by Bigelow, has three main attributes:

1. Task accomplishment: appropriate goals, undistorted communication, optimal power equalization based on problem-solving rather than position of authority;
2. Internal integration: resource utilization, cohesiveness, and morale;
3. Mutual adaptation of the organization and its environment: innovativeness, autonomy, adaptation, problem-solving adequacy.

Bigelow selected two junior high schools for his experiment; one served as the experimental school and it underwent an intensive organizational development program while the other school, serving as a control, did not. A pre post design was used to study the effects of the program in the classroom. The specific goals of the program were (Bigelow, 1970, paraphrased):

1. Increased openness and ease of interpersonal relationships;

2. A feeling that clearer and more effective staff communication had been achieved;
3. Increased willingness and skill in giving information about one's own behavior;
4. Increased awareness of interpersonal dynamics when they are taking place;
5. A widened and shared perception, by members of the staff, of other kinds of organizational patterns;
6. Skill in using a systematic group problem-solving procedure;
7. An increase in sharing initiative, that is, skill in helping a colleague who had enunciated an idea to develop it into a practical plan of action.

Before and after his training program, Bigelow used Flanders' Interaction Analysis, a lesson-plan analysis, and a classroom-life questionnaire to measure the effects of the training program. The results indicated that students' attitudes towards classmates improved, but their attitudes toward the overall class and teacher did not improve. No change was found at the control school. The Flanders' Interaction Analysis results showed a change toward a more open, inclusive, student-centered style of verbal behavior for the trained teachers but no change for the control school. Bigelow concluded that organizational development programs will have some effect in the classroom, but variables such as teachers' power, job security, percent of

time the participants work with superiors and peers, and relevance of training (seen by the teachers) should also be considered.

Bigelow has explored the concept of classroom climate and its relationship to organizational development programs in depth. Flanders (1970) quotes Bigelow:

The words classroom climate refer to generalized attitudes toward the teacher and the class that the pupils share in common in spite of individual differences. The development of these attitudes is an outgrowth of classroom social interaction.... These expectations color all aspects of classroom behavior, creating a social atmosphere or climate that appears to be fairly stable, once established. Thus, the word 'climate' is merely shorthand reference to those qualities that consistently predominate in most student-teacher contacts and contacts between pupils in the presence or absence of the teacher.... increased openness and ease of interpersonal relations and increased willingness and skill in giving information about one's own behavior would certainly have a positive effect on classroom climate.... A widened and shared perception of some new varieties of classroom organization would be a logical outcome of this type of two way communication.... Problem-solving skills would aid the student and the teacher in finding solutions using information from her expertise in classroom management and student feedback.... Therefore, if it can be shown that an organizational development program in the public schools does produce a more open, indirect style of teaching, and its concurrent open inclusive classroom climate, this would be a measure of change in the classroom level activity and of the increased effectiveness of the project school.

In general, these statements paralleled the intent of the Vernon School staff. However, further development of the evaluation procedure was necessary.

The Vernon staff concurred with Bigelow's hypothesis that organizational change should lead to classroom climate change. The staff felt that he had developed a useful evaluation design which should measure any behavior change. I agreed that the Flanders' Interaction Analysis of verbal behavior of the students and teachers was appropriate for the planned program's needs. I did not agree that the classroom-life questionnaire and the lesson-plan analysis were appropriate. A change in lesson-plans would be no indication of change in the students' or teachers' behavior, only what the teachers were reporting and planning to do. The classroom-life questionnaire was also inadequate. The attitude of the students to other students and to teachers was of concern; however, an attitude survey was inappropriate. The underlying assumptions of attitude tests (being able to read the questions, understanding the questions, answering truthfully, not being influenced to answer what students think they are supposed to answer, not being biased if the teacher had to read the questions) and finding a reliable instrument that the staff could later use unassisted, eliminated an attitude survey. The behavior of the students in the classroom is observable, measurable, and does not rely on students accurately or honestly reporting their feelings.

The Flanders' Interaction Analysis, as a measure of

change of verbal behavior, was recommended to the staff for part of the evaluation. The staff concurred; however, after examining the Flanders manual, I concluded (and the staff agreed) that they would not be able to use the instrument without experienced assistance. It was proposed that some simpler instrument should be used that correlated with the Flanders instrument. The staff could use this other instrument in the future; however, no other verbal instrument could be found so a search for a non-verbal instrument was conducted. The staff agreed that if verbal behavior changed (measured by Flanders) and non-verbal behavior changed in the same direction (measured by simple statistical procedures) then the staff could use the non-verbal instrument in the future. We assumed, based on Bigelow's and others' findings, that the behavior of the observed students would improve as a result of the organizational and communications programs implemented at the school. Chapter IV details the implemented programs.

II. NON-VERBAL BEHAVIOR OBSERVATIONS

The Flanders' Interaction Analysis instrument, described later in this thesis, has been found by many researchers to be highly reliable and has many well-established statistical procedures for the comparison of data and the recording of observations (Flanders, 1965, and 1970; Amidon, 1967). Bigelow (1970) stated that this type of

analysis related very well to the objectives of organizational and staff development programs. It provided a direct relationship between the skills learned during training and the categories of verbal behavior assessed by Flanders; however, it did not fulfill the requisites of the Vernon staff.

Many instruments were considered for the classroom behavior evaluation at Vernon. None was found appropriate for the objectives of the staff. The staff wanted specific categories of behavior to be recorded which were easily understood and observable by anyone on the staff. A short training time was desired so that teachers, aides, etc. could familiarize themselves with the instrument and be able to use it without having to be extensively trained or require assistance for interpretation. A statistical procedure that required little training or computational time was also necessary. The main objective was that little if any interpretation or judgment would be required by the observer while recording and interpreting behavior.

Four examples of instruments found in research papers are described here. See Simon and Boyer (1970) for an extensive presentation of instruments, all of which were reviewed for possible use or adaptation. No one instrument met all the staff requirements.

Madsen, et al., (1968) developed an instrument based on appropriate/inappropriate behaviors:

1. Inappropriate:

- a) gross motor: getting out of seat, running, etc.
- b) object noise: tapping feet, clapping, etc.;
- c) disturbances of other's property;
- d) contact: hitting, kicking, etc.;
- e) verbalization: any talking not permitted;
- f) turning around;
- g) other inappropriate behavior: ignoring teacher, daydreaming, etc.;
- h) mouthing objects;
- i) isolate play.

2. Appropriate behavior: time on-task, e.g., answers questions, listens, raises hand, works on assignment, etc.

This instrument was inadequate for several reasons: too much grouping of specific behaviors, especially of appropriate behaviors; turning around could be relevant and appropriate behavior; category "g" could include too many behaviors.

Kowatrakul (1959) developed an observation technique and a coding procedure for his research. The observation technique will be discussed in later chapters. His categories of behavior are:

- 1. Intent on on-going work: task oriented, performing assigned work;
- 2. Social work oriented: task oriented verbalization;

3. Social friendly: verbalization, non-task oriented;
4. Momentary withdrawal: not on task or interacting;
5. Intent on work in another academic area;
6. Intent on work in non-academic area;
7. Independent seat work;
8. Watching and listening: to appropriate task;
9. Discussion: three or more talking.

This categorical system was also found to be inadequate: these categories are only task oriented; tallies would report what was happening in general but not specifically what the student was doing; inappropriate behavior would mean the child was not on the task but would not tell what the child was doing instead of the task; the observer would have to be able to hear what the students were saying which is frequently not possible.

Craig and Holland (1970) developed an instrument to code specific non-attending behaviors but gave no attending behavioral categories:

1. Looking somewhere other than at teaching stimulus;
2. Fiddling;
3. Gross motor behavior;
4. Responses relating to teaching stimulus but inappropriate;
5. Temperamental behaviors.

Wandt and Ostereicher (1954) designed a fourteen-scale

system of categories. Both the teacher and the student could be rated on several categories (six categories for the students, eight for the teachers) on a nine-point scale; e.g., evidence of child-child cooperation: no cooperative work....almost all cooperative work, teacher's influence in decision-making: teacher makes all decisions....pupil makes most decisions.

This system required a great deal of judgment by the observer, an undesirable requirement. The categories also grouped all students together and did not identify specific behaviors. The rating system did not seem to be one which could be learned quickly and without assistance.

III. CONCLUSIONS

It was concluded that a new instrument would have to be designed that met the objectives of the Vernon staff. The instrument must be of little or no additional cost to the school; the staff must be able to use the instrument and interpret the results without the assistance of a trained researcher; little judgment must be required of the observer. Any instrument that did not meet these criteria would be of no lasting value to the staff.

The instrument I designed and field-tested must have the capacity to be used in more than one particular school to be valuable to me. A new instrument must have some gen-

eral usage to justify the effort required for validation and reliability studies. The teachers' objectives were the primary objectives in the designing of the new observation instrument; capacity for general adult-children interaction situations was the secondary objective.

CHAPTER III

DEVELOPMENT OF THE VERNON-NAY CATEGORY INSTRUMENT

I. METHODS-PART I

Initial Field-Tests

After reviewing the research literature, I arrived at general conclusions of what should be included in the category list. It was necessary, however, that the staff be involved in the selection of the categories as they would be using the instrument after the research project and I were finished.

One set of categories was classroom-tested before the final categories were determined. These preliminary categories were recommended by the staff as those behaviors with which they were most concerned. The initial categories were:

1. Participates in class activities/discussions:
student to student and student to teacher;
2. Antagonistic/belligerent toward the teacher or
other students;
3. Spends much time looking out of the window;
4. Disturbs other students;
5. Ignores requests;
6. Withdraws from class activities or leaves class;

7. Asks questions relevant to the subject;
8. Enthusiastic about the subject;
9. Number of times out of seat;
10. Number of times talking without permission.

The observers used the fourth-grade classrooms to determine if these categories were mutually exclusive and reliable. After three observation periods, it was apparent to the observers that these categories did not discriminate clearly; an inter-rater reliability test (see Chapter V) was calculated (0.340) and the necessity to revise the categories was obvious. Number six was too general; more categories were necessary; three and six, as well as four and two, overlapped; more positive categories were needed; a category for fighting was necessary as three fights occurred in two of the observation periods.

The staff and I defined a new set of categories based on my recommendations. The observers used these in three classrooms for a total of six hours. Inter-rater reliability coefficients were determined that indicated that these definitions were satisfactory; the coefficients were: 0.824, 0.863, 0.896.

The categories and definitions are titled the "Vernon Nay Category Instrument" (V.N.C.I.).

Categories Determined

The following are the behavioral categories used to record classroom behavior:

1. Enthusiastic about class activities: excited, waving hand for attention; "I know, I know;" gaining attention for the subject matter;
2. Helps the teacher or another student with a specific task, with teacher approval;
3. Asks for help from the teacher or another student, with teacher approval, about the subject;
4. Pays attention: doing assigned work without enthusiasm; conforming to expectations; working quietly alone; watching teacher or another student relevant to the assigned task;
5. Avoidance behavior: teacher-permitted behavior not related to the assigned task; sharpening pencil; leaving room with approval of the teacher, etc.
6. Seeks support from other students or staff: seeks attention by crying, clinging, showing affection; seeks sympathy;
7. Ignores or refuses requests from staff: ignores assigned tasks; playing with another student;
8. Leaves room without permission: all students are required to carry a hall pass;
9. Antagonistic or belligerent toward the teacher or

other students: verbal, gestural, pulling, pushing, kicking; not fighting;

10. Copies other student's work;

11. Withdraws from class activities: playing alone; no interactions; sleeps; stares; talks to self;

12. Participates in or starts a fight: actual blows; two students participating (distinguishes from antagonistic).

Each act was sub-coded as: In Seat/Out of Seat (coding: I/O); and Verbal/Non-Verbal (coding: V/NV).

Point-Time Sampling

The V.N.C.I. employs the point-time sampling technique developed by Kowatrakul (1959) in which a recorder watches each subject long enough to record one behavioral act, then observes the next subject. The recorder observes and records each subject in turn until all the subjects have been observed, and then starts over again. In this study, four students per class were observed (no set number is recommended by Kowatrakul). In a half-hour observation period, each student was observed an average of every seven seconds. Although only one student of the four is observed at one time, the observation period is so short that the procedure approaches simultaneous observation of the four students. A Sequence of Events Record was developed by me and my observers; (Appendix A).

Flanders' Interaction Analysis

The Flanders' Interaction Analysis (F.I.A.) is predicated on the assumption that verbal behavior of a student with a teacher is an adequate sample of either's total behavior in the classroom. The F.I.A. instrument is a systematic record of verbal behavior of the teacher and a student, one student at a time with the teacher. It is designed to distinguish between teacher behavior that decreases a student's freedom in the classroom (i.e., criticizing, justifying authority) by Direct Teacher Behavior, and that which increases a student's freedom of action (i.e., accepts or uses ideas of the student) by Indirect Teacher Behavior.

Five types (categories) of behavior are recorded: Indirect Teacher Behavior, Direct Teacher Behavior, Student Response, Student-Initiate, and Silence (or confusion). Indirect Teacher Behavior and Student-Initiate are the categories of behavior that Flanders assumes teachers strive for.

Flanders' categories are:

1. Teacher Talk:

a) Indirect Influence:

- i) accepts feelings: positive or negative feelings accepted and clarified in an unthreatening manner;

- ii) praises or encourages actions and talk;
- iii) accepts or uses ideas of students: developing student's ideas;
- iv) asks questions: about content or procedure;
- b) Direct Influence:
 - v) lecturing: facts or opinions about content or procedure;
 - vi) giving directions: commands, orders, students expected to comply;
 - vii) criticizing or justifying authority: intended to change a student's behavior; extreme self-reference.

2. Student Talk:

- viii) Response: responds to teacher;
- ix) Initiation: being called on if student expresses a desire to speak;
- x) Silence or Confusion: pauses or communication cannot be understood by observer.

The procedure for the F.I.A. instrument is that for each three seconds, the observer decides which of the stated categories best describes the communication pattern just completed. The observer writes down the behavior category number observed on a page that becomes a sequence of events record (no standard record necessary). The observer averages twenty observations per minute. Marginal notes may be

added as needed and if a major change in the classroom formation or communication pattern occurs, the observer draws a double line and notes the time. After the observation period is over, the observer writes a narrative summary of the class activities.

This procedure is used for the entire class as a whole but does not allow for observing the behavior of more than one student with the teacher at a time. The verbal and non-verbal behavior of the rest of the class is ignored.

Bigelow (1970) gives an extensive justification for using this instrument in observing the effects of organizational and staff development programs. In summary, three of the skills learned during the training of his experimental group (and the Vernon staff) were: paraphrasing, perception checking (relating how you think the other person is feeling), and behavior description (feedback to a person about what you have observed about that person). These three skills should increase the autonomy of the individual and the sense of equality when used honestly and with some level of expertise.

Bigelow states that one must understand what the other person wants and is trying to say before one can properly respond; feelings are always present and should be recognized and dealt with; paraphrasing and perception checking invite the student to respond to the teacher. When the teacher or another student disagrees with the behavior or

verbalization of another student, the student would not be criticized but the effects on the feelings of the class and teacher would be described by the teacher. Thus, the teacher is not demanding a change, only conveying information to the student about the effects of the student's behavior.

The Flanders analysis provides an assessment of teacher-student behavior in the classroom in relationship to the skills to be learned during the (proposed) Vernon staff training.

Conclusions

I decided that the two instruments would be used, concurrently, three times during the school year to determine whether behavior had changed as a result of the programs implemented by Stickel. Three observation periods, with two observers for every observation, would provide ample data to assess the inter-rater reliability and (concurrent) validity of the V.N.C.I. with the F.I.A.

It was apparent that the V.N.C.I. could be employed in many adult-children's group situations. "Copies" could be substituted for an appropriate category for each specific situation; minor definitional variations would probably be necessary for each environment in which the instrument was used. If reliability was significant (Chapter V), these categories together with the point-time sampling technique could provide a simple, easily learned assessment tool.

II. METHODS-PART II

Observer Training

Two adults were hired to be neutral observers in the classrooms. They were hired on the basis of past experiences with psychological research and classroom activities.

After defining the categories of the V.N.C.I., three days of preparation were needed to:

- a) develop tallying instruments that were facilitating to the observers (and, ultimately, for any staff);
- b) provide practice with the instruments;
- c) become familiar with the school's general operations.

The observers were trained with the final definitions and record sheets during two additional days in the fourth grade classrooms. This procedure allowed for a great deal of practice with the finalized instrument, provided examples to clarify categorical definitions, and enabled the observers to become familiar with general, Intermediate Team, procedures. The fourth-grade was selected as the practice rooms because the fifth- and sixth-grades had been selected as the observation rooms. The fourth-grade rooms were more similar, operationally, to the fifth and sixth because all three grades were administratively grouped as the Intermediate Team.

The observers were permitted to help clarify and

develop the final form of the instruments so that they felt they were an integral part of the evaluation research. They did not, however, know about the new staff programs or the expected students' behavior changes.

During the observation periods, it was essential that both of the observers were watching the same student at the same time; inter-rater reliability estimates would depend, in part, on simultaneous observation of the same student. It was imperative, however, that the observers disrupt the classes as little as possible; consequently, any device that required a sound as a signal, or any apparatus that might be of interest to the students (thus diverting their attention from their normal activities) could not be used. The observers positioned themselves across the classroom from each other so that they were in clear view of each other and all the subjects; they signalled each other by glancing at each other as they were ready to move on to the next student. This added time to each observation and was not desirable; however, under the circumstances, this was the best system that could be employed. A non-apparent signaling device should have been used but none was available at the time. A technique has been made public concerning this type of apparatus, after the Vernon project was already underway (Reynierse and Toevs, 1973).

Subjects

Ninety fifth- and sixth-grade students were selected from the total population of one hundred and forty students in those grades. The ninety students were chosen by random selection of six students from each of the fifteen different classes. Six students were chosen from each class with the assumption, based on attendance patterns, that at least four of the six would be present in an individual class on the day the class was to be observed. This allowed two students to be absent or out of the room; each class did have four of the original six present during the observation period. On several occasions, more than four of the six were present; in these instances, the first four students reselected (randomly) from the original six were observed. Thus four students were observed in each of the fifteen classes. Table I diagrams the schedule.

Five teachers in the fifth- and sixth-grades (three in the fifth, and two in the sixth) were observed while teaching three class subjects: Math (M), Reading (R), and Language Arts (LA). Thus five teachers were observed teaching three different classes, and during each class, four different students were observed. A total of 113 different students were observed (forty-seven were observed more than once). The school had a forty-three percent student-body turn-over during the school year.

TABLE I
OBSERVATION SCHEDULE

Teacher Class*		Obs.#1	Obs.#2	Obs.#3
1	R	4	4	4
	M	4	4	4
	LA	4	4	4
2	R	4	4	4
	M	4	4	4
	LA	4	4	4
3	R	4	4	4
	M	4	4	4
	LA	4	4	4
4	R	4	4	4
	M	4	4	4
	LA	4	4	4
5	R	4	4	4
	M	4	4	4
	LA	4	4	4
TOTAL		60	60	60

TOTAL = 180

* R = Reading
M = Math
LA = Language Arts

Every student, in the class to be observed, wore a name tag for the entire observation period. This reduced the possibility that students would know who was being observed. The observers identified the four selected students by the name tag. The teachers were not told who would be or was being observed at any time during the year.

Procedure

The observers used a Sequence of Events Record for each observed class; they used these in each of the fifteen different classes (Appendix A). Each of the behavioral acts recorded was coded with the In/Out of Seat letter and the Verbal/Non-Verbal letter under the appropriate heading at the top of the column. One student could exhibit only one category of behavior (sub-coded with one each of the I/O, V/NV letters) for each observation. Each record sheet included a description of the particular class being observed: observer, date, teacher, focus of observation (whole class, separate groups, independent activities), number of staff in the room, time-begin and time-end.

Recording of data was not begun until each teacher had made clear to the class what was expected of the students. This was necessary for the observers to be aware of the assigned task and, consequently, what was appropriate task behavior. The observers made no tallies when classes were in transition from one subject to another or when students went from one classroom to another.

Each of the classes had students from the fourth-, fifth-, and sixth-grades because the classes were grouped by subject-ability of the students. Only the selected fifth- and sixth-graders were observed.

The observers tried to reduce a halo effect by having as little contact as possible with the students during the class's observation period. They waited until each class had settled down and the students were not showing any curiosity about them before they began the observation periods. If a student did try to interact with the observers, the observers were polite but told the students they could not talk to them and please leave them alone so the observers could do their work. Several students were curious about the observers, but no problems arose that the observers could not deal with or could not ignore.

Since both the V.N.C.I. and the F.I.A. had to be used concurrently, the observers tallied the events in the classroom for the V.N.C.I. and tape-recorded the entire class proceedings. Data for the Flanders analysis was obtained by playing back the recordings and tallying the verbal behavior according to Flanders' instructions. The microphone was aimed at the teacher, and no major problems arose with using the two instruments simultaneously.

After each observation period, the observers counted the acts by category for each student and entered them on a summary sheet by class, teacher, date, etc. (Appendix B).

For the tape-recordings, the appropriate entry was determined, the data was tallied and similarly entered on summary sheets (Appendix C).

Expected Results

It was not possible to determine the effects of the planned programs on the students without developing a new observation instrument; consequently, the relationship of behavior determined by the Vernon-Nay Category Instrument and tallied by the Flanders' Interaction Analysis became as important a research issue as was whether either instrument recorded change in the behavior of the students. While the items in each instrument may not be compared item-by-item, the direction of change in positive and negative responses should be in the same direction. If a significant coefficient of correlation could be computed for the V.N.C.I. and the F.I.A., one measure of concurrent validity of the V.N.C.I. will have been established.

The staff decided that for the F.I.A., there should be an increase of teacher responses in the (grouped) categories of: accepts feelings, praises, accepts ideas, and asks questions (Indirect Teacher Behavior); Student-Initiate should increase; there should be a decrease in the number of teacher responses in the (grouped) categories of: lecturing, directing, and criticizing (Direct Teacher Behavior) and a decrease in Student-Responses.

For the V.N.C.I., there should be an increase of student responses in the (grouped) categories of: enthusiastic, helps, asks, and attending (Positive Behavior); there should be a decrease in the (grouped) categories of: avoidance, support, leave, ignore, antagonistic, copies, withdraws, and fights (Negative Behavior). For the categories of In/Out of Seat and Verbal/Non-Verbal, the teachers believed that as behavior improved (measured by either instrument) more students would be in their seats and quiet.

CHAPTER IV

STAFF IMPROVEMENT PROGRAMS

The programs described in this chapter are the programs proposed by Stickel and adopted by the staff. The behavior of the students was expected to improve (V.N.C.I. Positive and F.I.A.-Indirect, -Initiate) as a result of these programs.

Organizational Development (Shared Decision-Making)

The following is a detailed comparison of the shared decision-making model and the management by directives model of school administration. The description of the management by directives model is a generalization of traditional school management; every school has some variations on the general model but, in the main, the generalization holds true. The traditional, or management by directives model is detailed for the reader that does not have an education background; without both models being detailed, some readers would not appreciate the significant changes introduced into this school, accepted by the staff, and accepted by the principal.

In a directives model, the principal, or curriculum specialist informs the individual teacher of what their

objectives in the classroom should be, as directed by state, local administration and other superiors. The individual teacher's classroom objectives are incorporated into these directed objectives and it is up to the individual teachers to design their lessons and carry out these directed objectives. The teacher has little if any input into the curriculum specialist's office or to the building and district administration. In some schools there exists a team or unit formation of teachers organized around grade levels (kindergarten-third, third-fifth, etc.). These teams may have a team leader appointed by the principal to help teachers implement the directed objectives into the individual classrooms. The team leader is seen as an administrative superior to the classroom teacher and the classroom teacher is responsible directly to the team leader. Some principals have a cabinet system to help administer the school, but the principal retains the decision-making authority. Communication patterns are from the district administration to the building principal, from the principal to the assistants (if they exist) and then to the individual teacher. The teachers' objectives are subordinate to everyone else's in the administrative hierarchy and it is the teachers' responsibility to implement the directed objectives and their own objectives.

In the shared decision-making model, the staff was divided into four teams: three instructional and one support. Each instructional team (grades kindergarten-third, fourth-sixth, and seventh and eighth) had a team leader, certified

staff and/or a combination of intern-teachers, as well as the students in the grades of that team. This organization was the program proposed by Stickel to the staff. The staff agreed to implement his proposal.

The team leader was responsible directly to the team as well as the principal. The principal had hiring and firing authority over the team leaders, as well as any other staff member, but followed the wishes of a team in hiring and firing. It was the team leader's charge to facilitate the needs of the team on an organizational/regulatory level as well as to provide counseling and discipline for the students in their particular team; they were hired for a full year and evaluated at the end of the school year for rehiring by the team, as well as the principal; they acted as facilitators during team meetings. Each team had a stated set of goals and objectives relative to the needs, identified by the staff, for students in that team. Each individual teacher in a particular team had his or her objectives for the year. Teams met to organize these objectives into a set of team objectives so that teachers did not feel that they were working alone; they had the support of their team to implement individual and team objectives. The team could add objectives that were important for the whole team to carry out. All of these objectives were initiated and implemented on the team level; the school administration did not interfere with team objectives. The team, of course, took into account state and local administrative directives. This was facilitated by another team

called the Support Team.

The function of the Support Team was to facilitate and coordinate all the programs in the building. The Support Team consisted of all the team leaders, the curriculum consultant, building counselor, community agent, principal, reading specialist, administrative assistant, and the day care director. The Support Team was also concerned with the directives of the district administration and the needs of the school staff not in an instructional team (librarian, secretaries, custodians, cafeteria personnel, etc.).

The team leaders met periodically, at least once a week for an hour or more, with the staff of their teams to discuss team and individual objectives, regulatory demands, organizational needs, budget problems, personnel problems, etc. The team leaders presented problems or topics of discussion to the Support Team for help in solving a particular problem. No team was bound to accept suggestions from the Support Team.

In the shared decision-making model, the principal was a member of the Support Team; however, the principal was still responsible to the district administration for all activities in the building. In addition, the district expects the principal to carry out the directives which it imposes on all principals. That is why the traditional principal is able to give directives about building administration that are not debated. In this new model, the principal ranked only as an equal member of the Support Team, not the leader

of the team. The principal had an equal but not superior voice in the administration of the building. As the team leaders could present problems to the Support Team for assistance, the principal could do the same. Responsibilities to the district were presented at Support Team meetings and the Support Team was obliged to facilitate the principal's needs as it did the instructional teams' needs and the needs of other members of the staff. Everyone on the Support Team could contribute to the facilitation of the principal's needs. The principal looked to the Support Team for assistance in implementing district directives. Thus this model could not be organized and implemented without a principal that accepted this unique new role. The principal had to be committed to the team concept or there was no new model.

Other members of the staff, not associated with a team or on the Support Team, could present their problems directly to the Support Team or through a delegate. One of the Support Team could be asked to represent them, if they desired.

In this new decision-making model, communication was from any person or group to the Support Team and then from the Support Team to everyone in the school. This was accomplished by the distribution of the minutes of every meeting to all members of the staff. The team leaders worked with their teams to implement the suggestions from the Support Team.

The Support Team had a chairperson or leader just as the instructional teams did. The chairperson was elected by the

members of the Support Team to be responsible for meeting agendas, time-lines, place of meetings, etc. but was not in a position superior to any of the members of the Support Team. The Support Team also had a secretary to record and distribute minutes of the meetings.

In this new model, students with problems were referred directly to the team leader of that student's team. The team leaders had the option of referring students to the principal, administrative assistant, community agent, counselor, reading specialist, curriculum consultant, back to the classroom teacher, or outside the building to some other agency or program. The whole team had the responsibility for the student's behavior, so the resources of the team and everyone else in the district could be used for a particular student. On the other hand, in a directives model, students are referred at the teacher's option (usually to the principal). This places the responsibility and burden of resource knowledge on the teacher or principal rather than a team, team leader, and support staff.

When a teacher has a problem that needs to be resolved, this new model calls for the teacher to refer the problem to the team leader who may then carry the problem to another person or team. The team leader and that team are responsible for solving the problem, using any resources available and necessary. In the directives model, the teacher or team leader who has a problem may consult with anyone, either in or out of the building, that he or she believes might be able

to assist him or her. This places the responsibility on one person rather than an instructional team or, ultimately, the Support Team. In the shared decision-making model, the Support Team, rather than any individual, was ultimately responsible for all the problems of the school and the solutions for those problems.

This new model focused curriculum development on the introduction of new curricula methods and materials that were task oriented with specific goals in the affective and cognitive domains. Any curricula to be introduced was discussed at team meetings and each team could approve the new materials if the curricula fit into the team's objectives. Any team or teacher could ask for curriculum assistance when it saw the need but the curriculum consultant did not direct teams or teachers to use particular materials.

The directives model focuses on the responsibility of the teacher and the principal. The new model focused on the responsibility of the support and instructional teams.

Staff Development

Another proposal of Stickel's, which was accepted, was a workshop entitled "A Workshop Approach to Improving Organizational Effectiveness." Each team was informed of the availability of funds for its team and it was explained how each team could be involved. Each team understood that it did not have to participate but the program was available. If a team had decided not to participate, there would have been no punitive consequences. All three instructional teams

and the Support Team agreed to participate.

During the month of February, 1973, two of the teams completed the workshop. The Upper-Grade team completed the workshop during the summer. This team was not directly involved in the evaluation component of these programs and its late participation had no effects on the results of this research. The team leaders had additional exposure to the learned skills through summer workshops and involvement in training during the school year with their own team as well as with the Support Team. They were responsible for continuous support and growth of the staff in their teams. During team meetings, the team practiced the skills while conducting their usual business; the team leaders led the meetings.

During March, 1973, the Support Team took their training and concentrated on skill development for the rest of the year. No leader was responsible for the Support Teams' continuing growth. Both the Support Team and the instructional teams hired a consultant to lead the training workshops but no consultant was used during the rest of the year's team meetings.

Each team was required to set its own objectives for the year if it wished to take the workshop. While objectives for each team varied slightly, the general objectives of all the teams were:

1. Develop an effective organization to deal with children: organizational effectiveness must deal directly with the needs of children as a first priority; the needs of chil-

dren often make demands that exceed the skills of any one member of the staff, but, except on very rare occasions, would not exceed the combined skills of the staff;

2. Develop an effective communication base: a communication base is essential in order to move from a singular position of decision-making to a group decision-making process involving teachers, children, parents, other staff and administration;

3. Improve interpersonal relationships: a program for organizational effectiveness cannot be accomplished in a single workshop or a few faculty meetings; the growth and development of effective relationships will depend greatly on a continuous long-range program.

The goal of these objectives was to better meet the needs of children and staff at Vernon School.

As a result of the work done in the above areas, these secondary objectives could be accomplished:

1. Establishment of criteria for change mechanisms related to organizational development;

2. Involvement of staff in communication skills through needs assessment of the total school and classrooms;

3. Involvement of staff in school and classroom goal setting and implementation procedures, based upon the identified needs of children;

4. Evaluation of existing programs in light of those identified needs.

The training consisted of fifteen hours per person, on

a team basis, in the following skills:

1. Organizational skills:

- a) learn how to gather and organize data;
- b) learn to make decisions by consensus;
- c) learn to identify problems;
- d) learn group process techniques for different kinds of problems;
- e) learn to produce and apply criteria;
- f) learn the techniques for determining priorities;
- g) learn how to deal with forces that are working for and against the group;

2. Communication skills:

- a) to develop the skills of listening to others;
- b) learn to organize and articulate ideas;
- c) learn to read non-verbal clues of communication;
- d) learn how to separate ideas from personalities;
- e) provide feedback for clarification of ideas.

3. Interpersonal skills:

- a) development of a trust relationship with other adults;
- b) learn to cooperate with others;
- c) learn how to influence and be open to influence by others;
- d) learn to be open to feelings and sensitive to the feelings of others;
- e) assume a fair share of the responsibility for making decisions, planning and implementing

programs;

- f) learn to support other members in their efforts to carry out a task.

The success of this model depended on the upgrading of the process skills of the total staff in decision-making so that program changes would be made which would enable the team to reach its intended objectives. The instructional teams must have the skill to identify the needs of children, define its own objectives, develop and implement new curricula, maintain its own organization, use outside building programs and consultants, and communicate its needs to the Support Team (if necessary) through its team leader. Organizational skills will maintain the dynamic group process and allow the team to move from the level of an idea, to the constructing and implementing of a program change. The model presupposes that all staff members have and use enough of these skills to move from the regulatory level of administration to a level of performance where program change can take place.

The first observation period was in January, 1973, one month before the Intermediate Team was trained. The second observation period was in March, after the Support Team was trained. The third observation period was in May, two and a half months after the Intermediate Team's training.

CHAPTER V

RELIABILITY AND VALIDITY

I. RELIABILITY

Inter-rater Reliability

An inter-rater reliability statistic was used for the Vernon-Nay Category Instrument and for the Flanders' Interaction Analysis. Scott (1955) developed an inter-rater reliability statistic, " π ." The statistic assesses the extent of inter-observer agreement for a set of behavioral observations. It was designed specifically to classify a large number of observer responses into nominal-scale categories. Scott states that the categories must be mutually exclusive and observations must be duplicated on a random sample of all sets of responses possible.

Example I summarizes the responses of Observer "A" and Observer "B" for a random selection of students (15) during one observation period. The example is presented to describe the computation of " π " and because the column titled "Number Agree" will be referred to several times in the following chapters (Appendix D).

" π " may be calculated from raw data or from percentile data. In this example the statistics are not equal but are similar. In Table II and III, the lowest statistic

EXAMPLE I
SCOTT'S PI CALCULATED

CATE- GORY	OBS. A	OBS. B	%A	%B	% DIFF.	NO. AGREE	(AVER. %) ²
E	33	42	.048	.061	.013	30	.0030
H	5	5	.007	.007	.000	5	.0001
AS	10	9	.15	.013	.002	9	.0002
PA	246	233	.360	.341	.019	210	.1238
PE	22	22	.032	.032	.000	22	.0010
S	5	5	.007	.007	.000	5	.0001
I	280	289	.409	.423	.014	259	.1730
L	38	38	.056	.056	.000	38	.0031
AN	0	0	.000	.000	.000	0	.0000
C	9	9	.013	.013	.000	9	.0001
W	36	32	.053	.047	.006	32	.0025
F	0	0	.000	.000	.000	0	.0000
TOTAL	684	684	1.000	1.000	.054	619	.3069

$$Pi_{(raw)} = \frac{\frac{619}{684} - .3069}{100 - .3069} = 0.863$$

$$Pi_{(\% \text{ Diff.})} = \frac{(100 - .054) - .3069}{100 - .3069} = 0.922$$

computed is recorded.

The "Number Agree" column is calculated by "chi-square" tables of agree-agree, agree-disagree. These tables allow a detailed examination of what categories the observers confuse most often (Appendix D). Example I indicates that the observers most often disagreed about "enthusiastic," "paying attention," and "ignoring." These disagreements are consistent throughout all three of the observation periods. A discussion of these results is found in the section entitled "V.N.C.I.-II."

Scott states that an inter-rater coefficient of 0.850 is a reasonable level of performance.

Flanders' Interaction Analysis

As reported in Chapter II, the Flanders analysis has been reported to be highly reliable. For this research project, only one observer tallied the verbal behavior in each class (by playing back the tapes). This was necessary due to financial constraints. Periodic reliability checks were conducted for this one observer to insure the objectivity of the observer. Table II records the inter rater reliability coefficients for the Flanders analysis. All coefficients are calculated by Scott's "pi." Only nine checks were anticipated; however, a time lag occurred between tallying the second and third observation periods. Additional coefficients were computed to insure reliabil-

TABLE II
F.I.A. RELIABILITY COEFFICIENTS

Teacher Class		Obs.#1	Obs.#2	Obs.#3	
1	R M LA	.936	.947		
2	R M LA	.901		.925	
3	R M LA	.963			
4	R M LA		.942		.678* .891*
5	R M LA		.893	.897 .913 .900	.905* .889* .923*

(*Re-checks due to time lag)

ity. The coefficients ranged from 0.687 (a re-learning session) to 0.963.

Vernon-Nay Category Instrument

For this instrument, inter-rater reliability coefficients were calculated for every classroom observation period. Reliability was computed for the Verbal/Non-Verbal category (Table IIIa), for the In/Out of Seat category (Table IIIb), and for the Twelve (Summated) Behavior Categories (Table IIIc). For the V/NV category, coefficients ranged from 0.871 to 1.000; for the I/O category, coefficients ranged from 0.893 to 1.000; for the Twelve (Summated) Categories, coefficients ranged from 0.713 to 0.963.

No trend of reliability coefficients can be determined from the results presented in the tables. All the coefficients except one are over 0.80; one, of one hundred and thirty-five observation periods, is not significant. A total of six of the one hundred and thirty-five observations fall below 0.850, the limits of acceptability set by Scott.

The observers reported their opinions about the range of coefficients after the research was completed. They felt that the instructions given the class by the teacher was the variable that most influenced their agreement or disagreement about any single behavioral act. The teacher's instructions to the class determined what was appropriate

TABLE IIIa
VERBAL/NON-VERBAL RELIABILITY

Teacher Class		Obs.#1	Obs.#2	Obs.#3
1	R	.890	.931	.912
	M	.911	.989	.932
	LA	.889	.974	.946
2	R	.921	1.000	.992
	M	.975	.971	1.000
	LA	.896	1.000	.898
3	R	1.000	.981	.947
	M	.991	.907	.973
	LA	1.000	.949	1.000
4	R	.901	.979	.897
	M	.912	1.000	.948
	LA	1.000	1.000	.953
5	R	.871	1.000	.882
	M	.872	.874	.925
	LA	.873	.973	.950
*TOTALS		.899	.962	.947

*Totals are calculated by summing
across all classes and teachers for
one coefficient.

TABLE IIIb
IN/OUT OF SEAT RELIABILITY

Teacher Class		Obs.#1	Obs.#2	Obs.#3
1	R	.986	.971	.968
	M	.988	.967	.934
	LA	.967	1.000	.975
2	R	.999	.978	1.000
	M	.893	1.000	.985
	LA	.987	1.000	.951
3	R	1.000	1.000	.992
	M	1.000	.943	.932
	LA	1.000	.956	.987
4	R	1.000	1.000	1.000
	M	.998	1.000	1.000
	LA	1.000	1.000	.971
5	R	.993	.954	1.000
	M	.995	1.000	.987
	LA	.984	1.000	1.000
TOTAL		.986	.976	.986

TABLE IIIC

TWELVE (SUMMATED) CATEGORIES RELIABILITY

Teacher Class		Obs.#1	Obs.#2	Obs.#3
1	R	.945	.941	.961
	M	.912	.962	.871
	LA	.908	.877	.895
2	R	.933	.867	.890
	M	.881	.832	.843
	LA	.865	.857	.888
3	R	.817	.916	.951
	M	.713	.887	.953
	LA	.865	.857	.888
4	R	.943	.865	.876
	M	.876	.868	.823
	LA	.900	.963	.875
5	R	.861	.880	.879
	M	.922	.833	.935
	LA	.869	.886	.927
TOTALS		.879	.896	.895

or inappropriate, approved or unapproved behavior. When the observers did not understand the instructions, they disagreed about what was expected. They could not ask the teacher to repeat the instructions as that would introduce another variable in the classroom interactions.

A rank-correlation coefficient, r_s (Dixon and Massey, 1957) was computed between the class observation periods with the lowest reliability and the same class' percentile of agreement between the observers for the category "ignores." This was done as an estimate of the observers' opinions about the teachers' instructions. If the instructions were unclear to the observers, the students would not exhibit appropriate behavior (as recorded by each of the observers). The lowest ten inter-rater coefficients were ranked one through ten; the "ignores" percentile of agreements for each of the same ten classes was ranked one through ten; the coefficient or correlation was computed, $r_s=0.318$, and found not significant ($r_{s(0.95)}=0.549$, $N=10$). Other categories of negative behavior, for these same ten classes, were similarly correlated with the corresponding rank and also produced insignificant results. The observers' opinions remain opinions.

II. BEHAVIOR CHANGES

Winer (1971) discusses the Friedman test of the coefficient of ranked data. The data may be nonparametric for the Friedman test. The data for the F.I.A. and V.N.C.I. are not random samples of behavior with repeated measures. The data are three samples of behavior measured three different times with different subjects each time. This statistic allows coefficients of correlation to be calculated for the F.I.A. and V.N.C.I. if both instruments have data that can be ranked. The coefficient is an estimate of the significance of change of behavior during the three observation periods.

The F.I.A.'s statistical procedure for interpreting results (after a lengthy tallying and summarizing process) is a ratio of Indirect Teacher Behavior to Direct Teacher Behavior. Student behavior is not incorporated into the ratio. Table IV is a summary of Indirect/Direct, and the ratio of Student-Initiation in proportion to all Student Talk possible. The data in this table was summated across classes for each observation period; Table V shows the Indirect/Direct data ranked after summation. The (Friedman) "chi-squared ranks" test is not significant at the 0.950 level. The Student-Initiation ratios were similarly ranked and tested; they were not shown to be significant.

The power of the Friedman statistic (Blalock, 1972)

TABLE IV
I/D & S/I RATIOS

Teacher Class		Obs.#1		Obs.#2		Obs.#3	
		<u>I/D</u>	<u>S/I</u>	<u>I/D</u>	<u>S/I</u>	<u>I/D</u>	<u>S/I</u>
1	R	.210	.100	.230	.090	.290	.160
	M	.180	.110	.210	.070	.310	.170
	LA	.270	.050	.290	.040	.310	.110
2	R	.300	.120	.150	.080	.180	.090
	M	.230	.060	.210	.020	.170	.110
	LA	.340	.060	.310	.005	.190	.040
3	R	.260	.009	.140	.030	.210	.130
	M	.190	.040	.310	.040	.390	.030
	LA	.320	.095	.270	.075	.310	.040
4	R	.190	.055	.230	.030	.410	.070
	M	.290	.115	.190	.120	.310	.120
	LA	.110	.030	.120	.015	.310	.070
5	R	.290	.020	.200	.110	.210	.055
	M	.260	.035	.120	.070	.380	.105
	LA	.330	.050	.180	.055	.400	.090

TABLE V
I/D RATIOS RANKED

Teacher	Obs.#1		Obs.#2		Obs.#3	
	<u>I/D</u>	<u>Rank</u>	<u>I/D</u>	<u>Rank</u>	<u>I/D</u>	<u>Rank</u>
1	.223	1	.237	2	.305	3
2	.289	3	.231	2	.181	1
3	.257	3	.189	1	.243	2
4	.190	2	.187	1	.343	3
5	.293	3	.167	2	.131	1

Sum of Squares (time) = 1.600

Mean Square (w. teachers) = 1.000

Sum of Squares (w. time) = 10.000

$$x^2_{(\text{ranks})} = \frac{SS_t}{MS_{wt}} = 1.600$$

$$x^2_{(\text{ranks})} (.950) = 6.000 \quad (2 \text{ df})$$

$$\bar{r} = \frac{nW-1}{n-1}$$

$$W = \frac{SS_{\text{time}}}{SS_{\text{total}}}$$

$$\bar{r} = -0.050$$

is dependent on both the sample size and the number of categories used. It appears that for this research, either the sample size was not large enough (5) or the number of categories was not large enough (3) to reach significance, or the change of behavior was not significant.

Winer (1971) also presents the coefficient of concordance, $\bar{r}$, defined as the coefficient related to the average intercorrelation between the rankings assigned, or a rank difference correlation. A negative correlation, $\bar{r} = -0.05$, was found for the F.I.A. rankings (Table V).

The same procedure was used to test the significance of change of behavior measured by the V.N.C.I. These data were converted to several ratios of positive/negative/attending (neutral) behaviors: positive to all behavior; attending to all behavior; positive + attending to all behavior; Non-Verbal to all behavior; In Seat to all behavior, etc. None of the coefficients reached significance. The coefficient of concordance was also computed for each V.N.C.I. ratio and no significant results were reached.

III. VALIDITY

Content Validity

Nineteen teachers at Vernon School and seven from another school assisted in the preliminary definitions of the twelve categories of behavior possible for the V.N.C.I. All of these teachers approved the final definitions. Also, three psychologists (Thesis Committee) agreed that the twelve categories appeared to be mutually exclusive; tests of reliability would indicate additional content validity. The reliability coefficients computed for the V.N.C.I., for one hundred and thirty-five observation periods, reflected reasonable levels of agreement.

The definitions (categories) appear to be mutually exclusive and to demonstrate content validity.

Concurrent Validity

Coefficients of correlation between the V.N.C.I. and the F.I.A. were calculated using nonparametric statistics. The direction of behavior, over a period of five months, assessed by each instrument and correlated by grouping categories of behavior for each instrument, indicate the validity of the V.N.C.I. in relation to the F.I.A. The F.I.A. is based on the assumption that the verbal behavior of the students with the teacher is a representative sample of either's total classroom behavior. The V.N.C.I. measures

the behavior of the students not observable by the F.I.A. The correlation of direction of behavior change between the F.I.A. and the V.N.C.I. is an estimate of the (concurrent) validity of the V.N.C.I. if the basic assumption of the F.I.A. is accurate.

Rank-correlation coefficients (Dixon and Massey, 1957) were computed between the six V.N.C.I. ranked ratios with the F.I.A. ranked ratios. Each instruments' ratios per observation period were summated across classes and teacher, then each was ranked by observation periods. Table VI summarizes the results of the rank-correlations.

TABLE VI
F.I.A./V.N.C.I. RANK CORRELATIONS

VNCI Groups:	Pos.	Att.	P+A	Neg.	V/NV	I/O
Obs. Per.:						
1. Jan.	.850*	.300	.500	.900	.600	.650
2. Mar.	.950	.450	.800	.600	.550	.680
3. May	.900*	.850	.900	.300	.520	.630

* For all observation periods, except these two, Observer A and B ranked teachers/classes the same. For January and May, Observer A is shown; Observer B ranked these two periods higher (coefficients of 0.900, January, and 0.950, May).

These coefficients were calculated as indicators of the correlation between the F.I.A. Indirect/Direct ratios and categories of the V.N.C.I. for direction of change of behavior during the three observation periods. It was expected that the direction of change during the school year could be measured by both instruments. As Table VI demonstrates, the correlation of V.N.C.I.-Positive and the F.I.A. Indirect/Direct ratios is 0.850+ for each observation period. The other V.N.C.I. (grouped) categories have inconsistent correlations with the F.I.A. ratios. The V.N.C.I.-Positive (grouped) categories do correlate with the F.I.A. ratios of Indirect/Direct Teacher Behavior. While the correlations are high, a great deal of summing of data was necessary to calculate the correlations. This is the only computed statistic that indicates a significant relationship between the F.I.A. and any category(s) of the V.N.C.I.

CHAPTER VI

V.N.C.I.-II DEVELOPED

For the V.N.C.I., one significant result was found using three different statistical tests and calculating dozens of coefficients. The rank-correlation coefficient that was significant was encouraging; however, an examination of all the original categories, and specifically the categories the observers had the most frequent disagreement about, was conducted. From my Example I (page 46), the number of agreements for any category may be seen. The agree-disagree tables (Appendix D) used to calculate the number of agreements also showed what one observer tallied and what the other observer tallied for the same student (or class) at the same time.

Throughout most of the observations, the observers were unable to reliably distinguish between attending behavior, enthusiastic and ignoring behavior. Several suggestions were made by the observers:

1. Eliminate fighting; no fight occurred during the three observation periods;
2. Employ a signalling device if at all possible;
3. A more precise definition of ignoring is necessary;

4. Concentration was distinguishable from attending and should be added as a category;

5. Enthusiasm needed to be redefined, considering concentration;

6. Attending when the teacher is near you is different and distinguishable from attending when no one is paying attention to you;

7. The teachers and research staff recommended that the categories of V/NV and I/O be discontinued as no one could find any use or significance for them once they were tallied.

As a result of these suggestions and the analysis of the data for the V.N.C.I., the V.N.C.I. was revised. The revised categories are:

1. Concentration (inactive): sustained on-task; no communication with peers; self-directed;

2. Enthusiastic (active): about class activities; excited; waving hand for attention about the task;

3. Helping: teacher or another student with teacher approval;

4. Asks for help: from the teacher or another student (with teacher approval) about the task;

5. Task-oriented: conforming to expectations; not enthusiastic; teacher directed; peer communication about task; allows group interaction if on-task;

6. Support: seeks support from students or teacher; seeks attention or sympathy for non-task problem;

7. Ignoring (active): or refuses requests from staff; ignores assigned task; may appear to be on-task but in fact doing no work;

8. Avoidance (active): repeated teacher-approved behavior not related to task (e.g., sharpening pencil, getting a drink of water, etc.);

9. Copies other student's work;

10. Withdraws (inactive): from class activities; playing alone; not on-task; no interactions;

11. Antagonistic: or belligerent toward others (physical); fighting included as a footnote;

12. Disruptive (verbal): not on-task; bothering other students;

13. Leaves room without permission (or expelled, noted as footnote).

This revised set of categories constitutes the V.N.C.I.-II. The new categories were tested for inter-rater reliability. For fifteen observation periods, a combined pi coefficient of 0.957 was obtained. The pi coefficients for individual sessions ranged from 0.889 to 1.000. These new categories have better reliability than the original categories (Table VII).

TABLE VII
V.N.C.I.-II RELIABILITY

Teacher:	1	2	3	4	5
Class:					
R	.964	.953	.913	.982	.897
M	.949	.957	.947	1.000	.962
LA	.931	.889	.928	.976	.939

The F.I.A. I/D ratios were rank-correlated with the V.N.C.I.-II-Positive (grouped) categories after concurrent use in fifteen classrooms. A perfect (1.000) correlation was found between the two instruments' ratios for the desired categories (grouped). Other V.N.C.I.-II (grouped) categories were not significantly correlated with the F.I.A. ratios.

These categories, clearly, are more reliable and correlate more highly with the criterion measure, the F.I.A. I/D ratios. The V.N.C.I.-II appears to be a better instrument.

CHAPTER VII

CONCLUSIONS

I. RESEARCH RESULTS

Due to a lack of financial resources and an appropriate assessment instrument, the focus of this component of the total evaluation of the school had to be revised. It was financially impossible to compare Vernon School to another, similar, school to measure the effects of the planned programs with a control school. Also, it was not possible to determine the effects of the new programs on the students without developing a new observation instrument that could, eventually, be used by the school staff without professional assistance. The correlation of behavior determined by the Vernon-Nay Category Instrument and the Flanders' Interaction Analysis was as important a research issue as was whether either instrument recorded desired change in the behavior of the students. The categories in each instrument could not be compared item-by-item, but could be statistically correlated by comparing (grouped) desired and undesired behaviors as measured by each instrument.

It was anticipated that desired behavior would increase, measured by both instruments. This was not statis-

tically determined. The grouped categories of desired behavior measured by both instruments, were highly correlated when tested by ranking desired behavior by the time of the year. Desired behavior was measured, simultaneously, by both instruments as it occurred but it did not increase sufficiently during the year to be statistically significant.

The reliability of the V.N.C.I. was determined by an inter-rater coefficient of correlation. Reliability was highly significant throughout the year.

The statistical results of the V.N.C.I. were not satisfactory. The assumption made regarding change of behavior being measured by both instruments were either incorrect or could not be statistically determined. The staff, measured by attitude tests not discussed in this paper, were satisfied that they had learned a great deal of valuable information and acquired skills that would benefit them throughout their teaching careers. They were satisfied that Bigelow's theories of organizational change with communications workshop experience and classroom behavior change were correct. The fact that the instruments could not measure the change of behavior was of no concern to the staff. It did concern me.

The categories and definitions of the original instrument were revised (V.N.C.I.-II). These revisions resulted in higher reliability and a more positive correlation with the F.I.A. than the original instrument. A simple statis-

tical procedure (percentiles and bar-graphs) was taught to the staff that was interested in using the V.N.C.I.-II. It is being used by several staff members at Vernon School and has been experimented with in other schools and situations.

II. FUTURE RESEARCH

Two recommendations for continued research are apparent from the results of the statistical tests of the V.N.C.I.-II:

1. The instrument should be validated by other researchers in a variety of schools;
2. The instrument, possibly with minor variations, should be used in a variety of adult-children's group situations; the observation system and category concept should be easily translated to group homes, childrens' detention homes, various special classrooms, and perhaps in high schools and colleges; the general idea of defining categories of behavior with the people who want an evaluation instrument (used with the point-time sampling technique) should have many uses.

My third recommendation for additional research is to investigate the assumption that Flanders makes regarding his instrument; that is, that verbal behavior of the teacher with individual students is a sample of the behav-

ior of either the teacher's or the student's total classroom behavior. Were this true, significant (inverse) correlations would have been found when comparing the F.I.A.'s I/D ratios with negative behavior assessed by the V.N.C.I. or V.N.C.I.-II. Only the positive behaviors correlated significantly.

A SELECTED BIBLIOGRAPHY

- Amidon, E.J. Interaction Analysis: theory, research, and application. Reading, Mass.: Addison-Wesley Pub. Co. 1967.
- Anderson, G. & Wolberg, H.J. "Classroom Climate and Individual Learning." Journal of Educational Psychology, Jan. 1969, 59 (6), 414-419.
- Argyris, C. Integrating the Individual and the Organization. New York: John Wiley & Sons, Inc., 1964.
- Bennis, W. "Goals and Meta-Goals Lab. Training." NTL Human Relations Training News, 1962, 1 (3), 1-4.
- Berenson, D.H. "The Effects of Systematic Human Relations Training Upon Classroom Performance of Elementary School Teachers." Unpublished Paper, 1964.
- Bigelow, R. "Effect of Organizational Development on Classroom Climate." CASEA, Univ. of Ore., Aug. 1970.
- Blalock, H.M. Social Statistics. New York: McGraw-Hill, 1972.
- Bradley, J.V. Distribution-Free Statistical Tests. New Jersey: Prentice-Hall, Inc., 1968.
- Craig, H.B. & Holland, A.L. "Reinforcement of Visual Attending in Classrooms for Deaf Children." Journal of Applied Behavior Analysis, 1970, 3, 97-109.
- Dixon, W.J. & Massey, F.J. Introduction to Statistical Analysis. New York: McGraw-Hill, 1957.
- Flanders, N.A. Analyzing Teaching Behavior. Reading, Mass. Addison-Wesley, 1970.
- Flanders, N.A. Teacher Influence, Pupil Attitudes, and Achievement. Washington, D.C.: Dept. of HEW, U.S. Gov't Printing Office, 1965.
- Gallagher, J., et al. A System of Topic Classification Classroom Interaction Study. University of Illinois, 1966.
- Holland, J.G. "Technique for Behavioral Analysis of Human Observing." Science, 1957, 125, 348-350.
- Kerlinger, F.N. Foundations of Behavioral Research. New York: Holt, Rhinehart & Winston, Inc., 1964.

- Kowatrakul, S. "Some Behaviors of Elementary School Children Related to Classroom Activities and Subject Areas." Journal of Educational Psychology, 1959, 50, 121-128.
- Madsen, C.H., et al. "Rules, Praise, and Ignoring: Elements of Elementary Classroom Control." Journal of Applied Behavior Analysis, 1968, 1, 139-150
- Reynierse, J.H. & Toevs, J.T. "An Ideal Signal Generator for Time - Sampling Observation Procedures." Journal of the Psychonomics Society, Inc., 1973, 5 (1).
- Scott, W.A. "Reliability of Content Analysis: The Case of Nominal Scale Coding." Public Opinion Quarterly, 1955, Fall, 321-325.
- Schmuck, R. & Runkel, P. "Organizational Specialists in a School District." CASEA, 1971.
- Siegal, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.
- Simon, A. & Boyer, E.G. (editors). Mirrors for Behavior II: An Anthology of Observation Instruments. Philadelphia: Research for Better Schools, Inc., 1970.
- Spaulding, R.L. "An Introduction to the Uses of the Coping Analysis Schedule of Educational Settings (CASES) and the Spaulding Teacher Activity Rating Schedules (STARS). Durham, North Carolina, 1967 (mimeographed).
- Wandt, E. & Ostreicher, L.M. "Validity of Samples of Classroom Behavior for the Measurement of 'Social-Emotional' Climate." Psychological Monographs: General and Applied, 1954, 68 (5), 1-12.
- Ward, M.D. & Baker, B.L. "Reinforcement Therapy in the Classroom." Journal of Applied Behavior Analysis, 1968, 1, 323-328.
- Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.

LIST OF APPENDICES

APPENDIX	PAGE
A. V.N.C.I. Sequence of Events Record.....	72
B. V.N.C.I. Summary Sheet.....	73
C. F.I.A. Summary Sheet.....	74
D. Agree-Disagree Summary Tables.....	75

APPENDIX A

VNCI SEQUENCE OF EVENTS RECORD

Observer_____ Date_____ Teacher_____

Focus of Observation: Whole Class__ Groups__ Indep.
Act._____

No. Staff_____ Class_____

TIME BEGIN _____Code: In seat, Out of seat, Verbal, Non-Verbal

CATEGORIES:

Name	En	Hl	As	At	Av	Su	Ig	Le	An	Co	Wi	Fi
1	E	H	As	At	Av	S	I	L	An	C	W	F
2												
3												
4												
1	E	H	As	At	Av	S	I	L	An	C	W	F
2												
3												
4												
1	E	H	As	At	Av	S	I	L	An	C	W	F

.
 .
 .
 .

TIME END _____

APPENDIX B

VNCI SUMMARY SHEET

Observer _____ Date _____

Teacher _____ Class _____ No. Staff _____

Whole Class _____ Groups _____ Indep. Act. _____

Total Time _____

Subject	E	H	AS	AT	AV	S	I	L	AN	C	W	F	V	N	I	O
1.																
2.																
3.																
4.																
TOTAL																
TOTALS																

APPENDIX C
F.I.A. Summary Sheet

Date _____

Class _____

Teacher _____

SECOND

FIRST

	Acc.	Pr.	Use	Asks	Lect.	Dir.	Auth.	Resp.	Init.	Sil.
Accept										
Praise										
Uses										
Asks										
Lecture										
Direct										
Authority										
Response										
Initiate										
Silence										
TOTAL										

APPENDIX D
AGREE - DISAGREE SUMMARY TABLES

Child's Name _____

Teacher _____

Class _____

Observation # _____

		OBSERVER A												
		E	H	As	Pa	Av	S	I	L	An	C	W	F	TOTAL
OBSERVER B	E													
	H													
	As													
	Pa													
	Av													
	S													
	I													
	L													
	An													
	C													
	W													
	F													
TOTAL														