DOCUMENT RESUME

ED 126 152 TM 005 409

AUTHOR Lorentz, Jeffrey L.
TITLE The Design of a Study to Validate Teacher Competencies in Terms of Pupil Growth.
PUB DATE [Apr 76]
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976); For a related document, see TM 005 304

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Achievement Gains; Achievement Tests; Classroom Observation Techniques; Elementary Secondary Education; Performance Criteria; Statistical Analysis; Student Testing; *Teacher Behavior; *Teacher Certification; Teacher Evaluation; *Teacher Qualifications; *Teaching Skills; *Validity
IDENTIFIERS *Competency Based Teacher Certification; Georgia (Carroll County)

ABSTRACT

The major objective of the Carroll County Competency Based Teacher Certification (CBTC) Project is the development of a system for the recertification of teachers with classroom experience. A major goal is the development of a model for the identification and measurement of teacher competency areas; and the assessment of the extent to which these teacher behaviors (competencies) affect student outcomes, and the extent to which they relate to school goals and objectives. The certification process involves several stages including a screening procedure, classroom observations by trained observers in natural settings, and a study of certain relationships outside the classroom. Collection of student outcome data is used during the developmental stages, not as a basis for individual teacher certification. The overall plan involves several steps including (a) establishment of a competency list and related teacher behaviors and student outcomes, (b) selection of appropriate classroom behavior and student growth measures, (c) data collection in on-going classrooms, (d) reduction of observation data to competency-related scores, and (e) examining relationships between observed competencies and student growth. Replication of these steps successfully in additional classrooms will provide a validation of the findings and confirmation that some competencies were measurable and resulted in student growth. This in turn would provide a set of measurable teacher competencies that could be used as part of the certification process. (RC)

THE DESIGN OF A STUDY TO VALIDATE TEACHER COMPETENCIES IN TERMS OF PUPIL GROWTH

Jeffrey L. Lorentz

West Georgia College

A Paper presented at the 1976 Annual Meeting, American Educational Research Association, April 19-23, 1976, San Francisco, California.

INTRODUCTION

Encouraged by the federal support for the "Title" programs which proliferated during the mid-sixties, researchers have had the resources to investigate a broad spectrum of educational programs. This federal largess, however, was not without its drawbacks. Program guidelines, funding cycles and other considerations often proved so restrictive or unreasonable that the conclusions from such programs were insignificant, not replicative, or even worthless. Nor were the initiators of many of the studies, however, without fault. Bloom (1966) in his AERA presidential address contended that, while a 2,000 precent increase in federal funding for educational R & D provided a strong motive for studies of education, only one in 1,000 studies reported--3 per year--were crucial or significant. Consequently, while the era has been an exciting and productive one for research, it has often been frustrating as well.

Program evaluators and statistical and design consultants working with funded programs are often called upon mid-way into the studies. They are handicapped by insufficient prior planning, inadequate literature review, improper instrument selection, inadequate comparison groups, unreasonably small samples, and unrealistically short periods in which to test for treatment effect. Finally, the need to interrupt the study at crucial times to produce interim and final reports is especially trying--since the reports are often due before the final and all important data can be collected and analyzed.

The above is presented in contrast to the environment in which the present study has been conducted.

The Carroll County Competency Based Teacher Certification (CBTC) Project, conducted by the School of Education, West Georgia College, and Carroll County, Georgia, teachers, under an ESEA Title III Grant from the Division of Program and Staff Development, Georgia State Department of Education, is now in its third year of operation. The major objective of the project is the development of a system for the re-certification of experienced classroom teachers which is a potential alternative to the program-approved approach currently in use in Georgia.

The CBTC project has been characterized by a healthy freedom from many of the problems alluded to above. From the outset it has been proposed as a long-term study which would be likely to produce little immediate results. The project design has been flexible and subject to immediate modification as needed. The Georgia Department of Education, as the grant agency has provided minimal guidelines apart from a broad mandate to develop a competency based certification system, and has requested few formal reports.

This absence of unnecessary guidelines and constant interference by the grant agency along with a flexible design should not be viewed as a weakness. On the contrary, the project is established on a sound theoretical base (CBTC Librarian's Report, 1974). More important, it also has the advantage of regular, continuous contact with a team of consultants, each of whom is an outstanding authority on teacher effectiveness, educational research or statistical analysis. Finally, this freedom and

flexibility has been deliberately cultivated due to the prototypical nature of the project: one hoped to be worthy of replication.

The risks of such freedom are obvious and the investigators are apprised of these risks as well as the responsibility. The benefits to be gained, however, are felt to far outweigh the potential risks.

The Project Design

A major goal of the project is the development of a model for the identification and measurement of teacher competency areas; and the assessment of the extent to which these teacher behaviors (i.e. "competencies") affect student outcomes, and the extent to which they relate to school goals and objectives. Although the CBTC project is limited to Carroll County, Georgia, it is anticipated that the results of the study will be applicable in other school districts both within and outside the State.

The project design is based on the following theoretical framework: The certification process will involve several stages including a screening procedure, classroom observations by trained observers in natural settings, and a study of certain relationships outside the classroom. The collection of data relating to student outcomes will be used during the developmental stages, but not as a basis for individual teacher certification.

The overall plan which was established and followed during the past three years involved several steps including a) the establishment of a competency list and related teacher behaviors and student outcomes; b) the selection of appropriate measures of classroom behavior and student growth; c) the collection of data in on-going classrooms;

d) the reduction of observation data to competency-related scores; and e) examining anticipated relationships between observed competencies and student growth.

Subsequent to the successful completion of these preliminary steps, replication in additional classrooms would provide a validation of the findings and confirmation that some competencies were indeed measurable and their presence resulted in student growth. This in turn would provide the Department of Education with a preliminary set of measurable teacher behaviors which had been shown to result in student growth and which could then be used as part of the certification process.

METHOD

Introduction

The first year of the project (1973-74) was devoted to a review of the literature, site visits by project staff to similar projects and other developmental activities. Local teachers and administrators, assisted by project staff and consultants developed a set of generic teacher competency areas which were based in part on existing lists, but considered essential to all local teachers. As each competency was adopted as part of the definitive set, performance criteria were also listed. This added specificity to the definition of the competency and also provided guidelines for the selection of the measuring instruments which will be the heart of the competency-based certification system. Eleven generic competency areas and some 40 sub-competencies comprised the resulting list.

During the first year also, potential instruments were identified, additional staff were selected, and teachers were recruited to serve as subjects of observation. The second year (1974-75) was devoted to data collection and the present (1975-76) year given to additional data collection along with the analysis of the initial observation data. The remainder of the present paper is devoted to a discussion of these latter activities.

Subjects

The subjects of the study are the 60 teachers who participated in the CBTC Project during the 1974-75 school year and were selected as follows: All teachers in Carroll County were contacted in the Spring of 1974 and shown a slide/tape about the CBTC project and asked to volunteer as observers or observees. Three observers were selected and sixty teachers agreed to serve as subjects. Letters of agreement were signed outlining responsibilities of both parties. Summer workshops were conducted to familiarize participating teachers with all phases of study including competencies, observation instruments and student outcome measures.

These teachers, being volunteers, represented grades 1 - 12, a wide range of subject areas and varying levels of experience. Both male and female and black and white teachers are represented in the sample.

Student Achievement Measures

The Scott Foresman Initial Survey Test, Form A (IST) (Monroe, Manning, Wepman and Gibb, 1972) was used as the pretest (fall) measure in the first grade. Subtests include Reading and Math. Since the Scott Foresman

Reading program is used in Carroll County primary grades, Scott Foresman's IST is used as a placement test in all first grade classes. Its validity and reliability make the IST an appropriate instrument for use in the present study.

The selection of an end-of-year first grade measure was made after several different tests were tried with a small sample of students not in the project. The Comprehensive Test of Basic Skills, Level B, Form S (CTBS) (CTB/McGraw-Hill, 1974) was found to be the most appropriate in terms of difficulty. Students were able to respond to some of the items, but did not reach the ceiling. (Both floor and ceiling effects are problems with some tests in the lower elementary grades).

The CTBS includes a total Reading and total Math subtest. The reported validity and reliability of the test are satisfactory.

The Iowa Test of Basic Skills, Primary Battery, Level 7, Form 6 (ITBS-P) (Hieronymus, Lindquist and others, 1972), a widely used and well-known achievement test was used in the second grade.

The Iowa Tests of Basic Skills, Form 5 and 6, Levels Edition (ITBS) (A. N. Hieronymus and E. F. Lindquist, 1971) was used with the remainder of the classes in grades 3 through 8. The test includes Vocabulary, Language, Work Study Skills, Reading, and Math Subtests. As with the other measures, the test's reliability and validity are adequate. In addition, the State of Georgia makes use of the ITBS Battery in the 4th and 8th grades in the state-wide testing program.

The Tests of Academic Progress, Form S (TAP) (Houghton Mifflin, 1972) are designed to measure the extent to which the objectives of a basic area of high school instruction have been achieved. The six

subtests are: Social Studies, Composition, Science, Reading, Mathematics and Literature. The appropriate subject area test along with the Reading Test was used in the various high school classes. The TAP is used in the eleventh grade as part of the Georgia state-wide testing program. Validity and reliability of the instrument are satisfactory.

Student Self-Concept Measures

I Feel - Me Feel (IFMF) (Yeatts and Bentley, 1970) was selected for use with students in grades 1 - 3. This 40-item Likert-type scale (using five faces which range from happy to sad rather than numbers) is a self-concept measure appropriate for use with children at this level.

Scoring was based on locally-developed factor keys which were entitled: Academic, Self, Frustration, Femininity, Fun and Independence. Coefficient Alpha reliabilities for these scales ranged from .54 to .84 for the CBTC sample.

How I See Myself (HISM) (Gordon, 1968) was used as a measure of self-concept in grades four through eight (Elementary Form) and grades nine through twelve (Secondary Form). Both forms of the instrument were scored using four of Gordon's Keys entitled Teacher-School, Physical Appearance, Interpersonal Adequacy and Autonomy. Reliabilities for these scales, using CBTC data ranged from .64 to .83.

The Junior Index of Motivation (JIM) (Frymier, 1970) was also used with students in grades 9 - 12. This instrument is used to assess students' motivation toward school. Fifty of the 80 items are scored to produce a single index of motivation. Frymier (pp. 60-85) reports a number of studies in which the validity and reliability of the instrument were assessed.

Classroom Observation Instruments

The Coping Analysis Schedule for Educational Settings (CASES), (Spaulding, 1970) is a category system designed to measure pupil socialization and consists of 19 categories of student "coping" behaviors which are identified by descriptive statements such as "Aggressive Behavior," "Self-Directed Activity," "Observing Passively," etc. The CASES categories are arranged with the more active coping categories grouped at one end and the more passive categories at the other, but the numbers do not represent a scale. CASES has been developed over a ten year period as a result of more than 1,000 case studies in on-going public school classrooms and other educational settings. Relationships between CASES and student achievement have been established by McKinney and others (1975).

The Florida Classroom Climate and Control System (FLACCS) (Soar, Soar and Ragosta, 1971) examines the control tactics of teachers as well as their affective behavior. It includes items relating to the nature of classroom structure, teacher and student control strategies, and teacher and student affective behaviors, both positive and negative. In Follow Through Studies (Soar, 1973), data indicated that FLACCS discriminated significantly between programs and related to pupil growth.

The Observation Schedule and Record, Form 5, Verbal (OScAR 5V) (Medley, 1955) looks at the verbal behavior of teachers as perceived by students. It is based on 14 categories for teacher questions and statements and four categories for pupil--initiated utterances; by examining sequence it is possible to recognize some 600 transitions from one event to another. OScAR has been widely used for a number of years and has been the basis for several other instruments.

The Teacher Practices Observation Record (TPOR) (Brown, 1972) measures the congruency of observed teacher behavior with Dewey's philosophy of Experimentalism. Half of the 62 items describe behavior which reflects agreement with Experimentalism, while the other half reflect disagreement. All of the items reflect behavior which is found in public school settings.

Other Measures

Socioeconomic Status. An estimate of the socioeconomic status (SES) was obtained from each student based upon the occupation of the father (or the mother if she were the family's sole support). This occupational information was transformed to a 1-8 scale (1=high status, 8=low status) using a modified version of Warner, Meeker and Eells (1960) classification of occupations and levels. As an estimate of the reliability of the assignment of SES, two raters independently assigned SES scores to a small (N=77) sample of students using the procedure described above. The resulting intra-class R was .94. Students with missing SES were assigned the mean for their classroom, rounded to the nearest integer.

Interview. A fifty-five question interview, developed with the assistance of Selection Research, Incorporated, of Lincoln, Nebraska, was designed to be used as a potentially useful screening instrument. The questions are scored dichotomously by comparing each response with a criterion. Responses which paraphrase the criterion are scored as correct and those which do not reflect the criterion are not counted. Items are grouped into eleven sets of five corresponding to the eleven CBTC competency areas.

Data Collection

Student Outcomes. Achievement and self-concept measures were administered early in the Fall of 1974 in the classrooms of 60 Carroll County teachers, and posttests administered again in the classrooms of the same 60 teachers--all of whom were still in the project in May, 1975.

A group of specially--trained testers administered the tests in each classroom. Where necessary (i.e. in primary grades) the tests were administered to small groups of students and care was taken to insure that students understood the testing procedure, instructions, etc.

After the pretests had been administered, each of the approximately 1800 students in the 60 classes was observed using the CASES instrument. This task was completed in three weeks. Each child's record was scored to identify his coping style. At the end of the year, two observations were again recorded on every student in the 60 classrooms.

Process data. After the pretests had been administered and the CASES scored, six children with different predominant coping styles were identified in each classroom. These six students were used in subsequent observations using STARS. Six such observations were made in each classroom.

During the year, each teacher was visited three times, and three five-minute samples of behavior were recorded on each occasion on the TPOR. Each teacher was also visited on six other occasions on each of which the first five minutes of behavior were recorded on FLACCS, the next ten minutes on OScAR and the last five minutes on FLACCS.

Visits were scheduled in advance, but no attempt was made to pre-select the activities to be observed, since the intent was to record a representative sample of the behavior present in each classroom.

During the Fall, 1974 every teacher was interviewed with the CBTC interview schedule and has been interviewed again during the Fall, 1975.

Data Reduction

Initial Processing. Tests were hand or machine scored, keypunched and verified. Observation data were keypunched from the data collection forms and verified. Achievement tests were scored according to the keys provided by the various publishers. The self concept measures were reduced to subtest scores by summing various combinations of items. CASES frequencies for each student were scored using Spaulding's procedure (personal communication, 1974) to identify predominant coping styles for each student from Style A (aggressive, manipulative) to Style H (other-directed, task-oriented), as well as a Composite Score (styles A -H).

When CASES observations were made, observers distinguished two types of classroom settings: Teacher Directed (TD) and Program Directed (PD). There is evidence from Spaulding's studies and from preliminary CBTC Project analyses that changes in the classroom setting resulted in changes in the student's CASES scores. Scores were therefore obtained for students for the appropriate setting by combining data from both initial visits only when collected in the same setting. The same procedure was followed for the CASES data collected in final (post) visits. Thus for one student, one or two sets of pre-CASES scores could be obtained (e.g.: both observations occurred in TD Setting; both occurred in PD Setting; one occurred in PD Setting and one occurred in TD Setting), and one or two sets of post-CASES scores obtained.

In general, the following steps were taken to reduce the observational data on teachers.

STARS. The 475 cells in the student/teacher interaction matrix were weighted and averaged to obtain an overall score and portions of the matrix are combined to produce scores reflecting various teacher types. These 12 scores are listed in Table 1. One set of scores was obtained for each student observed during each observation period and for the overall class if the teacher interacted with students not specifically observed. These individual scores were averaged for all observations to provide class means to be used in subsequent analyses. The 475 cells were also retained as total frequencies across all observations for each teacher.

OScAR. The record of sequential teacher-student behaviors was scored to produce a 600 cell matrix of occurrences of interest (out of 9999 possible). These were weighted and reduced to percents to produce 22 keys, twelve of which relate to the Carroll County Competency list. These are also shown in Table 1. One set of 22 scores was obtained for each classroom observation period, and average scores for the 6 observation periods were then produced. The 600 cell matrix of occurrences for all observations was also retained for each teacher.

FLACCS and TPOR. These data were reduced from a series of items (FLACCS=186, TPOR=62) for six series of observations to competency scores for each teacher by Robert and Ruth Soar at the University of Florida. These competency scores are based on a priori competencies identified in the Carroll County CBTC Project. They were normalized for the total project group (N = 60) by converting items to T scores. By combining sets of these items, FLACCS yielded 28 scores and TPOR resulted in 21 scores. These are shown in Table 1. The 186 FLACCS item T scores and 62 TPOR T scores were also retained for each teacher.

TABLE 1

SUBSCORES FOR STARS, OScAR,
FLACCS, TPOR AND INTERVIEW

VARIABLE NUMBER	KEY	DESCRIPTION
		STARS:
1	S1	Overall composite
2	S2	The Story Teller
3	S3	The Boring Lecturer
4	S4	The Examiner
5	S5	The Entertainer
6	S6	The Controller
7	S7	The Counselor
8	S8	The Psuedo-Peer
9	S9	The Discovery Teacher
10	S10	The Socratic Teacher
11	S11	The Effective Manager
12	S12	The Effective Expository Teacher
		OScAR:
13	01	Deviant Behavior
14	02	Listening
15	03	Pupil Involvement in Subject Matter
16	04	Pupil Involvement in Planning
17	05	Clear Instructions
18	06	Pause for Student to Answer Questions after Instruction
19	07	Clear Explanations
20	08	Pause for Questions after Explanations
21	09	Pupil Speaks Freely
22	010	Self-control
23	011	Pupil Initiations
24	012	Question Difficulty
25	013	Teacher Statements
26	014	Confidence
27	015	Feedback from Teacher to Students
28	016	Rebuking Behavior
29	017	Managing Behavior
30	018	Lecturing
31	019	Questions Source Modified
32	020	Question Difficulty Modified
33	021	Question Quality
34	022	Permissiveness of Teachers

TABLE 1 (continued)

VARIABLE NUMBER	KEY	DESCRIPTION
		FLACCS:
35	F1	Student reduced deviant behavior
36	F2	Student better physical, mental health
37	F3	Student enjoys class, happy smiles
38	F4	Student on task, actively involved
39	F5	Student absence of withdrawn behavior
40	F6	Teacher pauses, elicits and responds to student questions before proceeding
41	F7	Same as F6 above but a summative measure
42	F8	Teacher uses a variety of methods, verbal, to deliver instructions
43	F9	Teacher uses a variety of methods, verbal, plus selected non-verbal, to deliver instructions
44	F10	Teacher uses a variety of methods--facial--gestural
45	F11	Teacher uses a variety of methods, verbal and non-verbal to deliver instructions
46	F12	Student self-directed to move toward task
47	F13	Teacher demonstrates proper listening skills
48	F14	Student able to speak freely
49	F15	Teacher utilizes non-verbal communication skills
50	F16	Student able to follow directions, on task
51	F17	Students feel free to interrupt presentations
52	F18	Student movement toward tasks
53	F19	Teacher provides feedback to pupil on his behavior
54	F20	Teacher maintains self-control in various classroom situations and interactions with student
55	F21	Teacher recognizes and treats individual student behavior
56	F22	Teacher uses more than one instructional activity simultaneously
57	F23	Teacher evidence of a personal one-to-one relationship with each student
58	F24	Student evidence of importance as class member--group involvement
59	F25	Teacher supportive classroom management
60	F26	Student evidence of enthusiasm
61	F27	Teacher listens to students and provides feedback
62	F28	Student high interest

TABLE 1 (continued)

VARIABLE NUMBER	KEY	DESCRIPTION
		TPOR:
63	T1	Teacher selects goals and objectives appropriate to pupil need
64	T2	Teacher matches student with appropriate materials
65	T3	Student on task actively involved
66	T4	Teacher gathers multi-level materials
67	T5	Teacher gives clear, explicit directions which are understood by students
68	T6	General task
69	T7	Teacher respects individual's right to speak
70	T8	Teacher helps pupil correct cognitive misperception
71	T9	Teacher accepts necessity of dealing with individual students on individual basis
72	T10	Student actively involved
73	T11	Teacher--evidence of praise and/or rewards in operation
74	T12	Teacher accepts and incorporates student ideas
75	T13	Student expresses ideas and opinions different to those of teacher and peers
76	T14	Teacher evidence of one-to-one counseling and absence of ...rejection (brush-off)
77	T15	Sum of T1, T2 and T3
78	T16	Sum of T12 for teacher and T13 for S
79	T17	Teacher control--general
80	T18	Pupil choice of subject matter
81	T19	Teacher choice of subject matter
82	T20	Expansive treatment of subject matter
83	T21	Restrictive treatment of subject matter
		INTERVIEW:
84	I1	Diagnosis
85	I2	Organization
86	I3	Teacher communication
87	I4	Coping
88	I5	Insightfulness
89	I6	Student--teacher communication
90	I7	Methodology
91	I8	Self-concept
92	I9	Emotional awareness
93	I10	Teacher growth
94	I11	Public awareness

INTERVIEW. Each interview was tape recorded and the tapes were transcribed. For each of the 55 questions, a correct response has been established (which may be paraphrased or stated in the subject's own words). Five items on each of the eleven CBTC competency areas are summed to produce a subscale score. These eleven scores, related to the eleven Carroll County competencies, are shown in Table 1.

The item-level data for STARS (475 interaction cells), OScAR (600 occurrences), FLACCS (186 item T scores) and TPOR (62 item T scores) were combined into a vector of 1322 variables for each of the 60 teachers.

Reduction of student data. The voluntary participation of teachers in the CBTC Project, lack of equivalent tests across grade levels, student attrition, and related problems, resulted in some significant complications in the analysis of the student data. In general, the following steps were taken, subsequent to scoring described above for achievement, self-concept and CASES.

On an instrument, by instrument basis (and within achievement tests, by subtest), in order to maintain as large a number of subjects as possible, student pretests and posttests were matched. Students with missing pretest or posttest were dropped for that portion of the analysis. In some instances, primary teachers taught one group of students Reading and exchanged them with another project teacher for Math. These students were assigned to the "correct" teacher for each subtest.

ITBS subtests for grades 3 - 8 were reduced to a common index by converting raw scores to grade equivalent scores using norms tables, and then subtracting students actual grade from his grade equivalent scores.

Using the BMDO4V Analysis of Covariance program (Dixon, 1974) adjusted posttest means (i.e., regressed gains) were computed for each variable. In every case, pretest and SES were used as covariates to adjust the posttest scores. Resulting means were transcribed and key-punched. The number of students and mean pretest and mean SES were also recorded.

CASES gains were computed separately for PD and for TD settings.

Analysis Procedures

During the 1st year of observation and data collection, some 50,000 punched cards of raw data were stored on disk files and a large number of variables were reduced to a smaller set of indices for each of the 60 classrooms. A major task during that year was the development of scoring routines and other analytical processes. Resultant statistical analyses were generally descriptive of the sample.

Due to the nature of the data collected (and supported by the advice of project consultants) the overall design for data analysis during the present year has been divided into three major phases. The first phase, the initial reduction of observation data to a set of "competency scores" and the computation of student gains, has been described above. The reduction of the observation data to competency measures related to the CBTC competency list was undertaken in response to a major project objective. The second phase is the examination of the relationships between the competency scores and student gains and the third is an examination of the observational data for patterns and an attempt to sort teachers into groups. The second and third phases have been undertaken more or less concurrently.

Competency scores and student gains have been related in a series of multiple regression analyses. Some of these are reported elsewhere at this AERA meeting in papers by Soar and Spaulding. Groups of teachers with similar competency scores (identified in the Phase 3 analyses) have been contrasted, using student gaines as dependent variables, in a series of discriminant analyses. The results of these analyses are discussed below.

Teacher groups. One major phase in the analysis of observational data has been an attempt to sort the teachers into groups with fairly homogeneous teaching styles. Teachers with similar profiles on the observational measures form similar clusters, or groups. To determine which teachers belong together, it is necessary to examine the profiles of all teachers and to find those with similar profiles. The number of groups can range from one (i.e., all teachers exhibit similar scores on all measures) to N, where N=the number of teachers (i.e., each teacher is considerably different from every other).

Several alternative statistical methodologies addressing this problem of grouping are available. Transpose Factor Analysis, or Q-Analysis, is one of these methods which was selected for the initial grouping.

The 19 x 25 (475 cell) matrix of STARS frequencies was summed across observations within each classroom, resulting in a vector of 475 "scores" for each of the 60 teachers. This total 60 x 475 matrix was transposed and factor analyzed using Guertin and Bailey's (1970) program ED501. Four factors were selected for rotation. These four factors represent an initial clustering of teachers according to STARS patterns.

The names of the teachers who had the highest loadings on each factor were examined by project staff to determine if similarities existed. It was apparent that neither grade level nor subject matter were held in common by the members of any group. In fact, the differences were more apparent than the similarities.

As a next step, the four groups were contrasted on the 12 a priori STARS keys through a discriminant analysis using Veldman's (1967) program DISCRIM. Results indicated that the groups were significantly different overall (Wilk's Lambda = 0.079; $F = 5.05$, $p < .0001$) and significantly different on six of the twelve scales. At this stage, the investigators felt that two conclusions were supported by the data. First, it was apparent that at least some of the a priori STARS keys adequately described teacher characteristics and, second, and more importantly, that the presence or absence of teacher characteristics, or styles of teaching, could be measured by direct observation in live settings.

The same four groupings (from the STARS Q-analysis) were subsequently contrasted on the 22 OScAR scores, the 28 FLACCS scores, the 21 TPOR scores, and the 11 Interview scales through additional discriminant analyses. Each of these confirmed that the groups differed significantly overall on each of these measures (except the interview), and that the groups were significantly different on several of the subscales on each instrument including 8 OScAR keys, 9 FLACCS keys, 5 TPOR keys and 1 Interview key.

These results confirmed that the groups exhibited stable patterns which could be observed directly through the use of different observation

schedules, by different observers and at different times. It also confirmed that some of the behavioral characteristics of teachers, or "competency scores" did exist in varying levels within the sample group. It was of particular significance to the investigators that these characteristics were exhibited by teachers working in natural settings and were not artificially induced by the creation of a "test" environment.

In a subsequent analysis, the transposed (475 x 60) STARS matrix was normalized (i.e., converted to T-scores) and then factor analyzed. Four factors were selected for rotation. These four rotated factors comprised different teachers. These groups also differed significantly overall and on 11 of the 12 STARS keys, 17 of 28 FLACCS keys, and 5 of 22 OScAR keys. This suggests that multiple analyses are necessary to adequately identify all the characteristics of the teacher groups.

Additional analyses have been undertaken. Two of these include the transposed factor analysis of the 1322 x 60 matrix (columns 1-475 = STARS, 476-1074 = OScAR, 1075-1136 = TPOR and 1137-1132 = FLACCS) using raw frequencies and frequencies normalized across all subjects.

Since Q-analysis, however, considers only the criterion of relative parallelness of teacher profiles in grouping and ignores overall levels of these profiles and relative scatter within profiles, two other methodologies are being employed to establish groupings based on STARS, OScAR, FLACCS and TPOR. The relative discrimination afforded by these groupings are being examined by means of Bottenburg and Ward's hierarchical analysis, (Veldman, 1966) and Guertin's distance analysis. (Guertin and Bailey, 1970)

Finally, in order to examine items in each instrument which discriminate significantly, each of the groups identified in the various

Q-analyses are contrasted on each of the 1322 items using a 1-way ANOVA. The F-ratio is not used as an inferential statistic (since many would be expected to be significant by chance) but rather as an index of the relative contribution of each variable to the overall discrimination among the groups.

RESULTS

As discussed above, transposed factor analysis, or Q-analysis, of the 60 x 475 cell STARS raw frequencies and the examination of the resulting four groups by discriminant analyses produced profiles which were interpretable and meaningful to the investigators. This led to further use of Q-analysis to identify teacher groups. A normalized 60 x 475 STARS matrix, a raw-frequency 60 x 1322 cell matrix, a normalized 60 x 1322 cell matrix, and a standard score (z-score) matrix were each subjected to Q-analysis and Varimax rotation of factors.

Following Q-analyses, the profiles of the teachers loading highest on the factors rotated (from two through nine, depending on analysis) were contrasted by discriminant analyses. STARS, FLACCS, OScAR and TPOR keys which discriminated significantly among the groups helped the investigators to label these groups.

The group profiles tend to be stable for the groups across instruments, but differences appear when the teachers are re-grouped on the basis of successive analysis. This suggests both the need for multiple instruments in assessing teacher behavior and the need to examine multiple dimensions of behavior.

Factor analysis of the transposed raw-frequency 60 x 1322 cell matrix and Varimax rotation of factors produced both two and three

interpretable groups of teachers. The results of discriminant analyses of the 3-group solution are discussed as an illustration of the findings.

Table 2 presents the Wilk's Lambda and F-ratios testing the significance of overall group differentiation on profiles from four instruments. These indicate that the groups are significantly different on these dimensions. Group centroids are plotted on the two discriminant axes in Figure 1 to illustrate the separation of the groups, and to illustrate that the several instruments reveal different patterns of discrimination.

Univariate F-tests revealed that the three groups differed significantly on 6 STARS keys, two FLACCS and one OScAR key, and nine TPOR keys. These are shown, along with group means in Table 3. Group 1 teachers are characterized in part by relatively high scores on the "Counselor" (S7), "Discovery" (S9), and "Student expresses different ideas" (T13) keys. The Group 2 teachers are partly characterized by higher scores on "Boring Lecturer" (S3) and "Teacher respects student's right to speak" (T7), and lower scores on "Controller" (S6), "Listening" (O2), "Student actively involved" (T10). Finally, some of the significant characteristics of the 3rd group are higher scores on "Examiner" (S4), "Student movement toward task" (F18), and "Teacher helps pupil correct cognitive misperception" (T8). Other differences can also be observed in Table 3.

Finally, to determine if the different groups produce significant differences in student gain, gain measures were contrasted in discriminant analyses. The student gain on ITBS subtests did not differ significantly. Neither did self concept (HISM) nor CASES setting 1 gains.

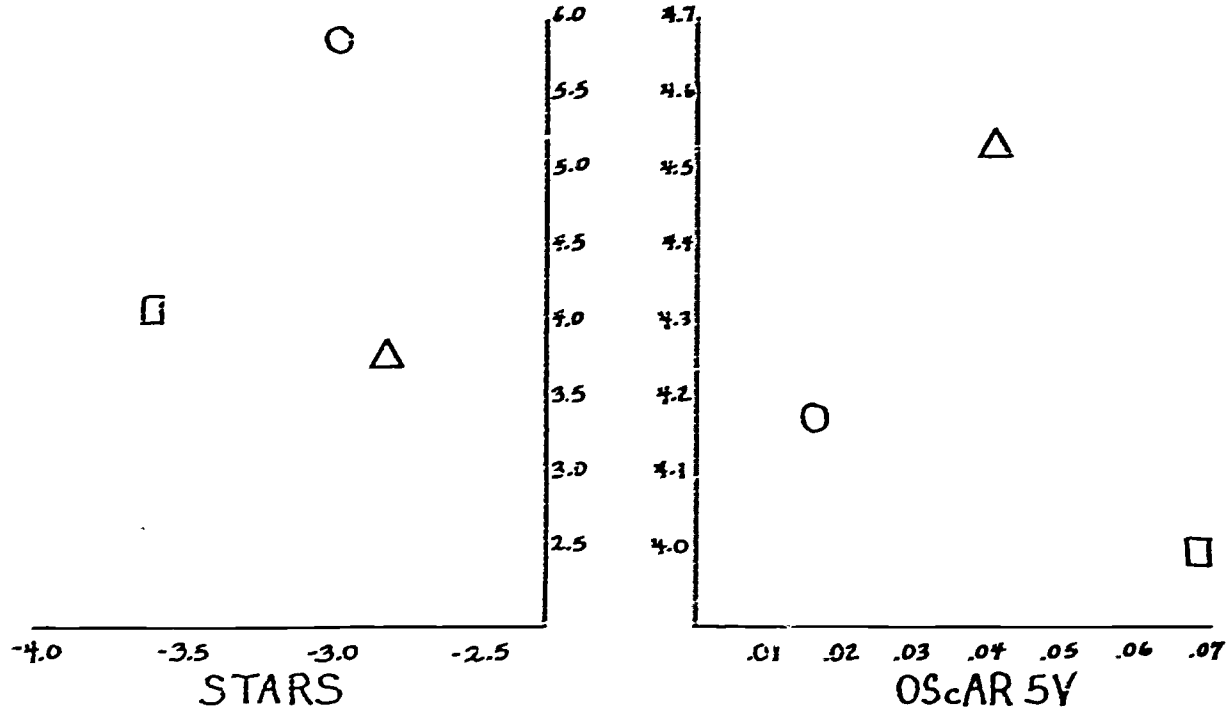
TABLE 2

RESULTS OF DISCRIMINANT ANALYSES

CONTRASTING 3 GROUPS ON STARS, FLACCS, OScAR & TPOR

INSTRUMENT	WILKS LAMBDA	df	F	p <
STARS	0.213	24, 92	4.464	0.0000
FLACCS	0.000	56, 60	1587.041	0.0000
OScAR	0.001	44, 72	41.479	0.0000
TPOR	0.044	42, 74	6.598	0.0000

Key
Δ = Group 1
O = Group 2
□ = Group 3



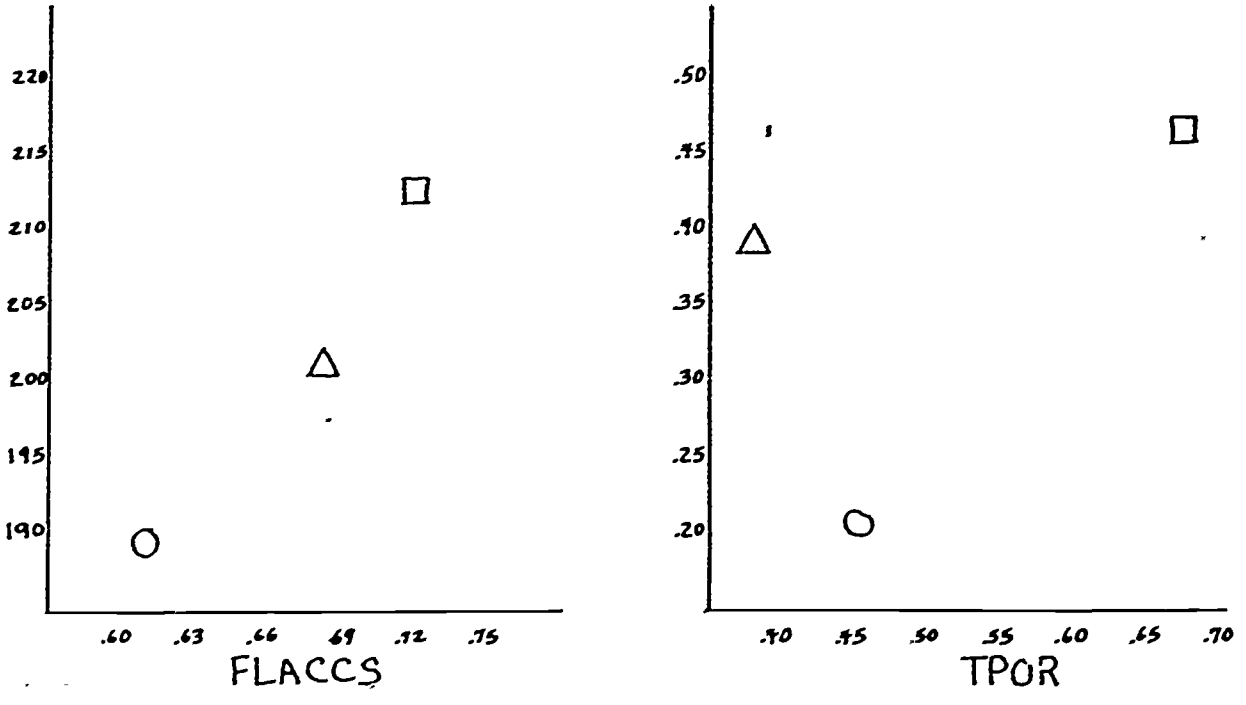


Fig.1 Group Centroids from 4 Discriminant Analyses.

TABLE 3

SIGNIFICANT ($p < .05$) CHARACTERISTICS

OF THREE GROUPS

FROM DISCRIMINANT ANALYSES

KEY*	GROUP MEANS 1	2	3	F (df = 2,57)	p
S3	16.8	32.2	18.4	28.32	.0000
S4	17.3	15.2	25.9	7.92	.0013
S5	10.5	14.9	7.2	6.34	.0036
S6	13.4	6.6	13.7	13.71	.0001
S7	3.3	1.2	1.2	3.68	.0305
S9	10.3	6.4	7.5	4.49	.0152
F18	255.5	230.9	264.6	5.06	.0096
F22	50.9	45.2	55.0	3.57	.0335
O2	4.8	2.9	6.4	4.04	.0222
T1	0.6	0.5	0.5	33.62	.0000
T2	0.5	0.7	0.5	76.61	.0000
T3	0.5	0.4	0.7	29.07	.0000
T7	16.8	32.2	18.4	28.32	.0000
T8	17.3	15.2	25.9	7.92	.0013
T9	10.5	14.9	7.2	6.34	.0036
T10	13.4	6.6	13.7	13.71	.0001
T11	3.3	1.2	1.2	3.68	.0305
T13	10.3	6.4	7.5	4.49	.0152

*Refer to Table 1 for description of keys.

However gains in CASES Setting 3 showed significant differences for Style E (in favor of Group 2) and for Style H (in favor of Group 1).

A difficulty in the separation of groups and subsequent examination of differences in student outcomes may be illustrated by the factor loadings on seven rotated factors from the analysis of the standardized matrix. Each of the seven factors had an approximately equal number of positive and negative high loadings. An identification of the names of the teachers with such loadings revealed groups who appeared (to CBTC observers) to lie at opposite ends of a continuum of behavior. Thus, each factor, or group, is in reality bi-polar and within group differences may counteract differences between groups.

SUMMARY AND CONCLUSIONS

The results of preliminary analyses suggest that systematic classroom observation using low-inference measures produces interpretable profiles of teacher behavior. Clear-cut relationships between teacher competencies and student outcomes have not yet been established; however, the data suggest that finer distinctions must be made among the groups before the competency-outcome relations may be uncovered.

The CBTC project has developed a prototype for the measurement of competencies as well as for the reduction of data related to teacher types. The observation instruments used in the study are worthy of retention. Each appears to identify unique characteristics of teachers, while at the same time complementing and supporting the others.

Further investigation of the same kind is needed using larger teacher groups at fewer grade levels. The reduction of N to 1 or 2

groups for some student measures made meaningful and valid analyses of some of the student data impossible.

Finally, it is apparent from the preliminary analyses that the most important relationships are probably non-linear. The task of discovering which of the many possible interactions among the teacher variables contribute to differences in student growth will be most difficult.

It is clear that the road to an empirically--founded competency based certification system will be a long one, filled with obstacles and detours. However, the goal is one which can be reached and the CBTC project has taken the first steps down that road.

REFERENCES

Bloom, B. S. "Twenty-five Year of Educational Research," American Educational Research Journal, 1966, 3, 211-218.

Brown, B. B. "Experimentalism in Teaching Practice," Journal of Research and Development in Education, 1970, 4, 14-22.

Carroll County Competency Based Teacher Certification Project. Librarian's Report. Mimeographed, 1974.

Comprehensive Tests of Basic Skills. Monterey, California: McGraw-Hill, 1974.

Dixon, W. J. (Ed.) Biomedical Computer Programs (3rd Ed.). Berkeley: University of California Press, 1974.

Frymier, J. R. "Development and Validation of a Motivation Index," Theory into Practice, 1970, 9 (1), 56-88.

Gordon, I. J. A Test Manual for the How I See Myself Scale. Gainesville: Florida Educational Research and Development Council, 1968.

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

Hieronymus, A. N., & Lindquist, E. F. Iowa Tests of Basic Skills. Boston: Houghton Mifflin Company, 1971.

Hieronymus, A. N., Lindquist, E. F., & Others. Iowa Tests of Basic Skills. Boston: Houghton Mifflin Company, 1972.

McKinney, J. D., Mason, J., Perkerson, K., & Clifford M. "Relationship Between Classroom Behavior and Academic Achievement." Journal of Educational Psychology, 1975, 67 (2), 198-203.

Medley, D. M. Observation Schedule and Record, Form 5, Verbal. Mimeographed, 1973.

Monroe, M., Manning, J. C., Wepman, J. M., & Gibb, G. E. The Scott, Foresman Initial Survey Test. Glenview, Illinois: Scott, Foresman and Company, 1972.

Soar, R. S. Final Report--Follow Through Classroom Process Measurement and Pupil Growth, (1970-71). Gainesville: Institute For Development of Human Resources, University of Florida, 1973.

Soar, R. S., Soar, R. M., & Ragosta, M. Florida Climate and Control System: Observer's Manual. Gainesville: Institute for Development of Human Resources, University of Florida, 1971.

Spaulding, R. L. Classroom Behavior Analysis and Treatment, San Jose: Author, 1970.

Tests of Academic Progress. Boston: Houghton Mifflin, 1972.

Veldman, D. J. Fortran Programming for the Behavioral Sciences. New York: Holt, Rinehart and Winston, Inc., 1967.

Warner, W. L., Meeker, M., & Eells, K. Social Class in America, New York: Harper, 1960.

Yeatts, P. P., & Bentley, E. L., Jr. I Feel--We Feel Self Concept Appraisal. Authors, 1970.