DOCUMENT RESUME

ED 135 761 SP 010 815

AUTHOR Mahan, James M.; Harste, Jerome C.
TITLE Professional Judgment as a Criterion Variable in Pre-Service Teacher Education Research.
PUB DATE 7 Apr 77
NOTE 13p.; Paper presented at the American Educational Research Association Annual Meeting (New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS College Supervisors; *Educational Research; *Effective Teaching; *Participant Satisfaction; Peer Evaluation; *Preservice Education; Self Evaluation; Student Evaluation; *Student Teachers; Student Teaching; *Success Factors; Teacher Education
IDENTIFIERS Purdue Student Teacher Opinionnaire

ABSTRACT

Data collected and analyzed in this study over a three-year period lend credence to the notion that professional judgment is a reliable and stable criterion variable for teacher education research. Participants included 209 student teachers, 131 classroom teachers, and 27 university teacher trainers. Student teachers were ranked according to their teaching effectiveness by their supervising teachers, their university methods instructors and supervisors, their fellow student teachers, and by a self-rating procedure. Fourteen of fifteen correlations among the various effectiveness rankings were positive and statistically significant at the .01 level. The remaining correlation was positive and significant at the .04 level. Student teachers ranked as the most effective professionally tended to be the most satisfied with their field-based teacher preparation program. Student teacher satisfaction with their training program was measured through use of the Purdue Student Teacher Opinionaire (PSTO). Supervising teachers expressed their perception of student teacher satisfaction by completing a Modified PSTO. There was considerable congruence between student teacher expressed satisfaction and teacher perception of the degree of student teacher satisfaction. (Author/MM)

Session 21.12

AERA Presentation

N.Y., N.Y.: April 7, 1977





# Professional Judgment as a Criterion Variable in Pre-Service Teacher Education Research

James M. Mahan, Indiana University
Jerome C. Harste, Indiana University

Turner (1975) lists professional judgment as a criterion variable related to teacher work success. This study focuses upon the issue of longitudinal reliability and stability of professional judgment as a criterion variable. Specifically, the purpose of the study was to annually collect and interpret data pertinent to the following questions of central importance to evaluators of teacher education outcomes:

(1) Do student teachers as a group, the individual student teacher, university methods and supervisory personnel, and supervising teachers agree in their professional judgments of the most effective student teachers?

(2) Are student teacher evaluations of the student teaching experience, as measured by the Purdue Student Teacher Opinionaire (PSTO) and by a preparatory program evaluation instrument, related to the effectiveness rankings of student teachers by peers, by self, by university personnel, and by supervising teachers?

(3) Are student teacher responses on selected subsets of the PSTO (i.e., Rapport with Students) related to one or more of the independent rankings of professional effectiveness?

(4) What degree of congruence is there between student teacher evaluation of the student teaching experience via the PSTO and supervising teacher evaluation of the same experience via a modified PSTO?

(5) How stable are these relationships over three annual replications?

## RATIONALE

Turner (1975) has suggested that the aim of research in teacher education is to optimize that portion of teacher work success attributable to teacher preparation. Teacher work success is seen as a complex notion involving at least three classes of criterion variables: (1) student attainment, (2) professional judgment, and (3) student judgment. In further explanation of the criterion variable of professional judgment, Turner states that the teacher's professional image or stature, as judged by school and college supervisors, peers, and school administrators is an important dimension of teacher success since virtually all teachers are employed in some kind of institutional setting or formal work organization. Included in a teacher's professional image may be ability to work with peers, relation to parents and other community members, perceived motives, knowledgeableness, leadership qualities, perceived skill, dependability, dedication and so on. One difficulty in optimizing teacher success is that the criteria mentioned by Turner above, and all of which are important, may not invariably correlate positively with each other. In certain instances inverse correlations may appear.

Decreases in pupil dependency or autonomy may accompany increases in student achievement, or high levels of student achievement may be associated with low ratings of instruction. If the criterion which represents teacher work success is indeed as heterogeneous as suggested above, a distinct optimization problem is posed. It is difficult to determine the means by which, and those points up to which, one set of criterion variables can be increased without inducing a decrease in another set of criterion variables. For example, to what point can student attainment be increased without decreasing teacher ability to function in school work settings? This same difficulty may well exist within criterion variable classes, i.e., supervising teacher perceptions of teacher success may well vary from university personnel perceptions of the phenomenon, and both may vary from peer perceptions.

Given these possibilities, the purpose of this study is seen as one designed to explore the nature of the relationships which exist within the criterion variable of professional judgment. Conflicting qualitative perceptions of teacher performance are often assumed to be an issue confronting teacher education evaluators. This study provides needed data relative to that issue.

## PROCEDURES AND INSTRUMENTATION

Four independent rankings of student teacher effectiveness were obtained at the end of the 1972-73, 1973-74, and 1974-75 academic years. Data were obtained from all participants involved in a year-long, field-based teacher education program and included 46 classroom teachers, 85 student teachers, and 10 university methods instructors and supervisory specialists in 1972-73; 45 classroom teachers, 70 student teachers, and 8 university personnel in 1973-74; and 40 classroom teachers, 54 student teachers, and 9 university personnel in 1974-75.

Program Satisfaction Measures. Student teacher satisfaction with the student teaching program was determined through administration of two different instruments, the Purdue Student Teacher Opinionaire (PSTO) and a questionnaire developed to provide specific information about key dimensions of the field-based training program, the Program Evaluation Scale (PES). The PSTO has high validity and reliability coefficients (Bentley and Price, 1969) and has been employed frequently to measure satisfaction with student teaching conditions in both traditional and innovative programs. The PES has been employed successfully since 1973 in the evaluation of the referent teacher training program (Mahan and Loadman, 1973). A Modified PSTO, revised to measure the supervising teacher's own opinions of the student teacher's satisfaction with the conditions and relationships encountered during student teaching, was administered to each supervising teacher. By using the Modified PSTO, a comparison of student teacher satisfaction with the classroom teacher's perception of that satisfaction was possible, i.e., where the student teacher reacted to a statement like, "My relations with other teachers in the building were ...," the classroom teacher reacted to the reworded statement, "The relations of my student teacher with other teachers in the building were ...."

Student teachers made a second expression of satisfaction with their training program via the Program Evaluation Scale (PES). A ten point continuum ranging from "1" (no contribution) to "10" (great contribution) was utilized to assess the degree to which the student teachers felt that eight different aspects of the program contributed to their professional development. The aspects were:

- supervisory assistance from classroom teachers
- instructional preparation received from methods professors
- assistance received from supervision specialists
- supervisory assistance received from methods professors
- advantages of being placed in 2 or more schools
- advantages of working with 2 or more groups of pupils
- assistance received from peers as a result of the "cluster" approach to student teaching
- preparatory impact of the total program

A student teacher could record summated evaluative scores as low as 8 or as high as 80 on the PES and as low as 100 or as high as 400 on the PSTO. Those with the highest PES scores and highest PSTO scores were considered to be the most satisfied with their year-long student teaching program as measured by the specific instrument. A highest third, middle third, and bottom third of student teachers was identified on the basis of summated PES scores and again on the basis of the summated PSTO scores.

Student Teacher Effectiveness Measures. Professional effectiveness rankings collected included (1) student teacher rankings of peers, (2) classroom teacher rankings of student teachers, (3) university staff rankings of student teachers, and (4) student teacher self rankings.

Teacher effectiveness was defined by each group operating independently but with the input of the other professional groups involved in the study. This was accomplished in the following manner:

(1) A whole group meeting involving classroom teachers, student teachers, and university methods persons was held to discuss various criteria for judging teacher effectiveness. All participants understood that teacher effectiveness was a complex notion oftentimes defying easy explanation and that, therefore, each group could independently select the final criteria they would use.

(2) Following this meeting, each group met independently to arrive on the criteria it would employ in judging teacher effectiveness.

Using the procedural preferences expressed by student teachers relative to the Student Teacher Ranking by Peers, the names of all student teachers were placed on a sheet of paper. This listing was given to the student teachers along with instructions to rank order the top six student teachers with respect to their classroom teaching effectiveness. All responses were anonymous. The sum of ranks for each student teacher, based upon the responses of peers, was computed. Student teachers with the highest sum of ranks were considered to be the most effective professionally. Those student teachers receiving few or no votes were considered to be least effective professionally. The student teachers were then divided into a highest third, middle third, and bottom third on a basis of professional effectiveness for data analysis purposes.

Using the procedural preferences which were expressed by classroom teachers, rankings of student teachers were obtained by having each classroom teacher rate each of his/her final two student teachers on 10 criteria for professional effectiveness. A five point scale ranging from "1" (some) to "5" (outstanding) was employed. The criteria were:

- effective classroom management
- creation of a motivating learning environment
- preparation of sound lesson plans
- verbal reinforcement of pupils
- sensitivity to individual needs of pupil
- reliability
- acceptance of classroom responsibility
- readiness for daily teaching duties
- enthusiasm for teaching
- mastery of general teaching skills

Each student teacher could earn a summated score of from 10 to 50. Student teachers with the highest total score on the 10 criteria were considered to be the most effective professionally. The student teachers were then divided into a highest third, middle third, and bottom third for data analysis purposes.

University Staff Rankings of Student Teachers were obtained by having each staff member rate every student teacher on five criteria. A five point scale ranging from "1" (unsatisfactory) to "5" (outstanding) was employed. The criteria were: rapport with children, planning for instruction, initiative and creativity, classroom management, and implementation of teaching plans and activities. Classroom teachers and staff members in a series of conferences concerning the evaluation of student teachers had agreed that these five criteria included the ten criteria utilized by the classroom teachers. Student teachers with the highest summated score on the five criteria were considered to be the most effective professionally. Those with the lowest scores were considered to be the least effective professionally. For data analysis purposes, the student teachers were again divided into a highest, middle, and bottom third.

Student Teacher Self Rankings were obtained by having each student teacher rate his or her self on the same ten criteria used by the classroom teacher in rating the student teachers. The same five point scale was employed. Each student teacher could award himself from 10 to 50 points. Student teachers receiving the most points from themselves were considered the most effective professionally. Classification into a highest, middle, and bottom third was done.

Data Analysis Procedures. Spearman's Coefficient of Rank Correlation (Siegel, 1956) was used to examine relationships among the four rankings of student teacher effectiveness, the PSTO, and the PES. Similarities and differences between pre-service and in-service teacher responses on the PSTO, Modified PSTO, and selected subscales of the PSTO were examined through use of arithmetic means and graphs.

## RESULTS AND DISCUSSION

Fourteen of the fifteen correlations displayed in Table I are positive and statistically significant at the .01 level. Furthermore, these rank-order correlations indicate a tendency of the underlying ranks to relate in a monotone-increasing manner (Hays, 1963, p. 655). For example, low rankings by student teachers are associated with correspondingly low ranks by supervising teachers and so on. These significant correlations occurring over three successive years in an on-going, field-based teacher preparation program suggest considerable

TABLE I

Spearman Rank-Order Correlations for Four Rankings of Student Teacher Effectiveness

| Year | N | Variable Identity | Var. 1 | Var. 2 | Var. 3 | Var. 4 |
|---|---|---|---|---|---|---|
| 1974-75 | 54 | #1 | 1.00 | | | |
| 1973-74 | 70 | Ranking of Student | 1.00 | | | |
| 1972-73 | 85 | Teachers by Peers | 1.00 | | | |
| | | #2 | | | | |
| 1974-75 | 40 | Ranking of Student | .25 .04 | 1.00 | | |
| 1973-74 | 45 | Teachers by Class- | .53 .01 | 1.00 | | |
| 1972-73 | 46 | room Teachers | .34 .01 | 1.00 | | |
| | | #3 | | | | |
| 1974-75 | 9 | Ranking of Student | .67 .01 | .53 .01 | 1.00 | |
| 1973-74 | 8 | Teachers by Univer- | .60 .01 | .56 .01 | 1.00 | |
| 1972-73 | 10 | sity Staff | .55 .01 | .38 .01 | 1.00 | |
| | | #4 | | | | |
| 1974-75 | 54 | Ranking of Student | .47 .01 | .41 .01 | .36 .01 | 1.00 |
| 1973-74 | 70 | Teachers by Self | .54 .01 | .49 .01 | .41 .01 | 1.00 |
| 1972-73 | 85 | | *... | *... | | |

*No Student Teacher Self Ratings were obtained at the end of the 1972-73 academic year.

agreement among individual student teachers, student teachers as a group, university personnel, and classroom teachers concerning who is, and who is not, a professionally effective student teacher. Table I data indicate that the various program participants tended to select about the same subgroup of students as "most effective" and the same subgroup as "least effective." These perceptions of effectiveness were stable from year-to-year. One explanation for this finding in this program is that university personnel and classroom teachers probably had reached many common educational expectations and operational agreements in the three years this program existed prior to this study. These common expectations, transmitted to student teachers at the beginning of the academic year, no doubt led to the various types of program participants sharing similar perceptions of teaching success. If this explanation is accepted then one might expect that if a similar study had been conducted in the first year of implementation of the same preparation program it might have revealed more conflicting views of teaching success.

An alternate explanation of the data is that professional judgment of teacher effectiveness is a relatively stable and reliable criterion variable despite the profession's seeming inability to operationally define it. Many of us in field-based programs have always suspected this, feeling intuitive judgments about teacher effectiveness were better than the performance profiles which our data provided.

Our own experience operating field-based programs allows us to conclude that almost everyone can agree informally on who are the really "good" teachers and who are the really "weak" ones. The problem is not agreement, but definition of the items among which agreement is found.

TABLE II

Spearman Rank-Order Correlations Between Four Rankings of Student Teacher Effectiveness and Satisfaction With the Student Teaching Experience as Measured by the PSTO and PES

| Variable Identity | Year | Satisfaction With Field Experience | |
|---|---|---|---|
| | | PSTO Total Score | PES Total Score |
| Ranking of Student Teachers by Peers | 1974-75 | .00 | .30 .02 |
| | 1973-74 | .14 | .31 .01 |
| | 1972-73 | -.07 | * ... |
| Ranking of Student Teachers by Classroom Teachers | 1974-75 | .32 .01 | .22 .06 |
| | 1973-74 | .44 .01 | .47 .01 |
| | 1972-73 | -.42 .01 | * ... |
| Ranking of Student Teachers by University Staff | 1974-75 | .08 | |
| | 1973-74 | .28 .02 | .37 .01 |
| | 1972-73 | -.18 .06 | * ... |
| Rankings of Student Teachers by Self | 1974-75 | .31 .02 | .43 .01 |
| | 1973-74 | .33 .01 | .42 .01 |
| | 1972-73 | * ... | * ... |

*The PES device and the Student Teacher Self Rating Device were not utilized at the end of the 1972-73 academic year.

Table II indicates that student teachers who are perceived as highly effective professionally in a field-based preparation program tend to perceive the program itself as highly effective when they are evaluating the program using the PES scale, an instrument designed to specifically address the major characteristics of the program. Relative to the standardized PSTO scale (where some subscales deal with forces beyond university influence like school curriculum and materials, community support, and principal's leadership) there is, due to the 1972-73 data, a less stable tendency for the "best student teachers" to be the "most satisfied" with their year-long student teaching experience. No clear pattern emerged between a student teacher's satisfaction with the student teaching experience as measured by the PSTO and the effectiveness ranking attributed to that student teacher by peers. One explanation for the negative correlations between PSTO satisfaction and rankings in 1972-73 is that in years of teacher unrest or budget reduction the

components of the PSTO that focus upon community, administration, and public school curriculum can evoke negative responses from the "best" student teachers. Effective teachers may, for example, be the most sensitive to breakdowns in either the community's or the university's provision of quality educational services and may assign low evaluations to the responsible party. Turbulence characterized the 1972-73 year for participants in this study. School board decisions to form middle schools and to change attendance areas, reassign elementary pupils, and transfer teachers upset both supervising teachers and student teachers. Negative correlations between professional effectiveness and satisfaction with the program as measured by the PSTO occur only in 1972-73. It is unfortunate that no PES data was collected in 1972-73. If PES data had been obtained, and if it had correlated positively with effectiveness rankings, the argument that community variables beyond university staff control strongly effected student teacher satisfaction would be strengthened.

TAELE III

Spearman Rank-Order Correlations Between Scores on Selected Subsets of the PSTO and the Four Rankings of Student Teacher Effectiveness

| PSTO Factor | Year | Ranking: By Classroom Teacher | | By Self Rating | | By Peers | | By University Staff | |
|---|---|---|---|---|---|---|---|---|---|
| F1 Rapport with Classroom Teacher | 1974-75 | .15 | | -.06 | | -.03 | | .03 | |
| | 1973-74 | .27 | .02 | .18 | | .15 | | .32 | .01 |
| F3 Teaching as a Profession | 1974-75 | .18 | | *.32 | .01 | .02 | | .18 | |
| | 1973-74 | .46 | .01 | .46 | .01 | .11 | | .11 | |
| F4 Rapport with University Supervisor | 1974-75 | .10 | | *.26 | .03 | .31 | .02 | .19 | |
| | 1973-74 | .44 | .01 | .23 | .03 | .15 | | .33 | .01 |
| F7 Rapport with Students | 1974-75 | .16 | | .20 | .08 | .08 | | .04 | |
| | 1973-74 | .36 | .01 | .27 | .02 | .20 | .05 | .33 | .01 |
| F10 Professional Preparation | 1974-75 | .06 | | *.50 | .01 | *.28 | .03 | *.24 | .05 |
| | 1973-74 | .51 | .01 | .38 | .01 | .29 | .01 | .30 | .01 |

*Most stable relationships across years.

Table III reveals that students who thought most highly of their own professional effectiveness, as measured by PSTO, were also the most satisfied with teaching as a profession, with their relationships with the university supervision specialists, and with the quality of methods instruction (professional preparation) received from the methods professors. The same students were also rather positive about

their rapport with their elementary school pupils. It is also evident that satisfaction with professional preparation (positive feelings about methods instruction) was positively associated with high effectiveness ratings. Although correlations fluctuate from year-to-year, the general trend is for the student teachers judged most effective to be the student teachers most satisfied with the five selected aspects of their teacher preparation program as measured by the PSTO.

Graph I presents a comparison between student teacher satisfaction with student teaching as measured by the PSTO and supervising teacher perception of that student teacher satisfaction as measured by a Modified PSTO.

Subscales of these instruments are:

Factor 1. Rapport with Supervising Teacher
Factor 2. Rapport with Principal
Factor 3. Teaching as a Profession
Factor 4. Rapport with University Supervisor
Factor 5. Community Support of Education
Factor 6. Student Teacher Load
Factor 7. Rapport with Students
Factor 8. Rapport with Other Teachers
Factor 9. Satisfaction with Housing
Factor 10. Professional Preparation
Factor 11. School Facilities
Factor 12. Curriculum Issues

If one defines .5 (one-sixth of the range) as a difference of practical concern, there are few differences among group means that are significant. Generally, classroom teachers rather accurately perceived the extent of student teacher satisfaction with the year-long field experience.

The 1974-1975 teachers and student teachers had significant differences regarding Factor 2, Rapport with Principal, and Factor 8, Rapport with Other Teachers. The 1972-1973 student teacher/supervising teacher groups revealed differences of .5 or greater on two factors, also: Factor 5, Community Support of Education, and Factor 8, Rapport with Other Teachers. The classroom teachers in these cases thought satisfaction was greater than it actually was. The disparity between student teacher and classroom teacher on the issue of satisfaction with community support, to some degree lends credence to the earlier interpretation given to the negative correlations displayed in Table II for 1972-73.

GRAPH I PSTO FACTOR SCORES COMPILED BY 1972-73 AND 1974-75 STUDENT TEACHERS (STs) AND SUPERVISING TEACHERS (CTs)

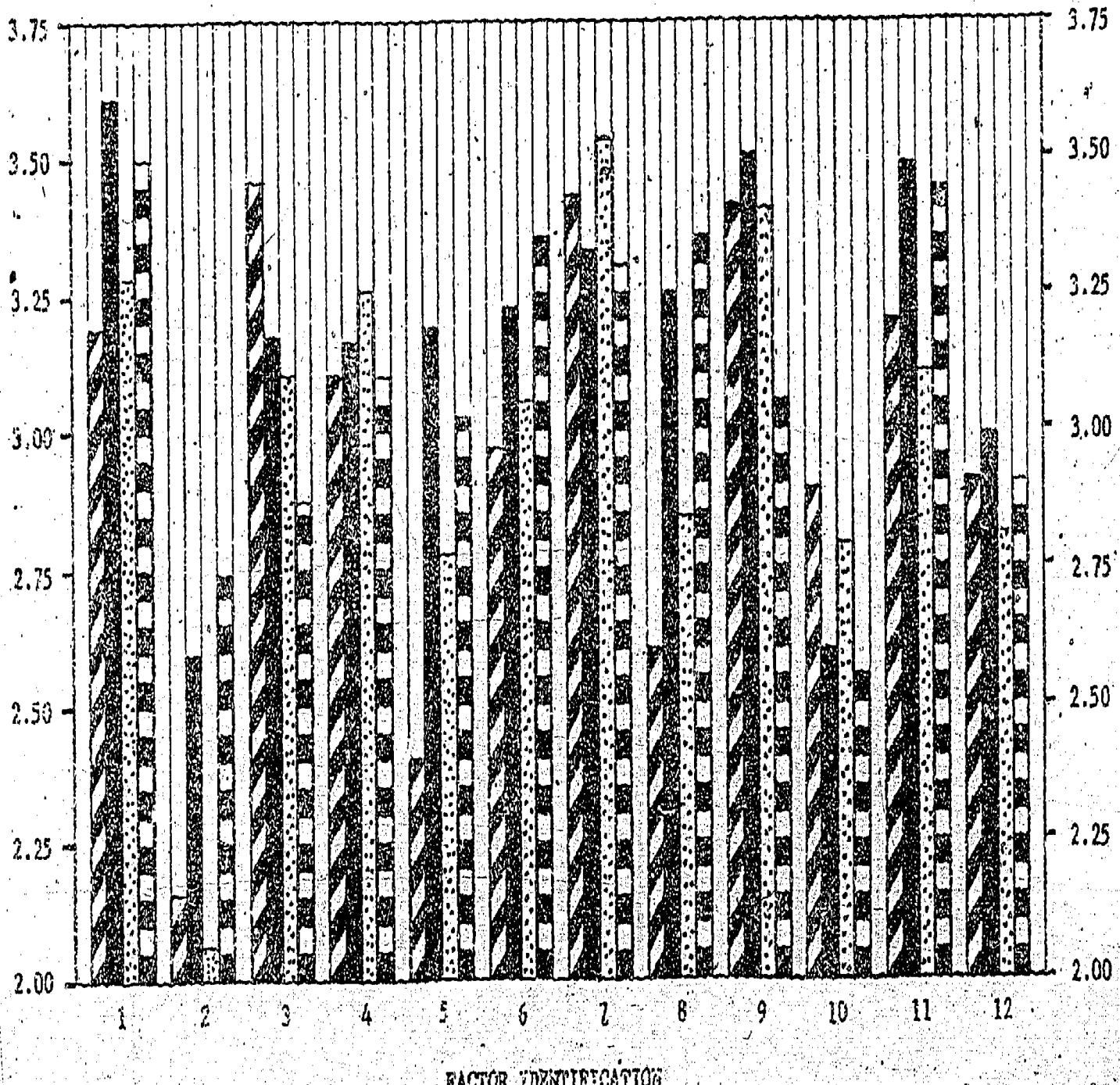
= 1972-73 STs (N=65) $\bar{X}$ = 2.98 S.D. = .37

= 1972-73 CTs (N=35) $\bar{X}$ = 3.17 S.D. = .34

= a range of .5 PSTO scale points (defined as a difference of practical significance in this study)

= 1974-75 STs (N=57) $\bar{X}$ = 3.01 S.D. = .30

= 1974-75 CTs (N=40) $\bar{X}$ = 3.10 S.D. = .31




FACTOR MEAN SCORES

FACTOR IDENTIFICATION

Then too, there seems to be a consistent tendency for classroom teachers to feel student teachers have better rapport with the remainder of the faculty than actually exists. This phenomenon occurred in both years. One explanation for this finding is that student teachers probably do not share with the current supervising teacher all the delicate details of a strained relationship with a different teacher elsewhere in the building. Another explanation is that some supervising teachers who may be virtual "loners" in terms of school socializing may not even imagine that a student teacher desires open and friendly interaction with many classroom teachers -- not just one. Furthermore, almost all the classrooms in this program were "self-contained." It may be that inservice teachers are forgetting the limitations the self-contained classroom places on intra-faculty communication, grade-to-grade sharing, team teaching, or even on a common coffee break time. Idealistic beginners on the other hand, desire and anticipate these close collaborative conditions and are disappointed if they do not exist.

A third and final explanation for the variation in the mean scores between groups on Factor 5 and Factor 8 is that these subscores are the least stable of all scores on the PSTO. This explanation has some basis in that an examination of the PSTO reveals these subscales are made up of only 5 and 6 items respectively. While the least exciting, this explanation may well be the most plausible. If accepted, Graph 1 provides even stronger evidence supporting the realiability and stability hypothesis being tested in this paper.

## CONCLUSIONS

Data collected and analyzed in this study over a three year period lend credence to the notion that professional judgment is a reliable criterion variable for teacher education research. Whether or not professional judgment is as reliable as student attainment and student judgment is an empirical issue which needs to be studied. Similarly, whether or not professional judgment is a valid measure of teacher work success, or whether or not professional judgment is as valid a measure as are student judgment and student attainment, were issues not researched, and clearly ought to be. Asking questions of this sort may, however, be misleading, for somehow the mere act of suggesting that the validity of professional judgment be studied elevates, and with no clear support, the stature of student attainment and student judgment as valid criterion variables for judging teacher work success. At this point there is little to suggest that one criterion variable is more valid than another. If anything, the record favors professional judgment. The long time favorite, student attainment, must certainly be questioned given the reliability problems identified by Soar and Soar (1975). Then too, student attainment logically suffers in that it is a product measure and teaching is, we all agree, a rather complex and dynamic process. Professional judgment as a criterion variable both permits and allows this important dimension of teaching to enter into the evaluation process. And this we believe is a very positive and desirable feature of professional judgment as a criterion variable in teacher education research. This point, however, probably needs to be restated. In essence, what we are saying is that teaching is a complex and dynamic process, and while researchers in the field would readily agree, their most commonly employed methods and criterion variables treat it as a static set of skills and as a product rather than a process. Consider, for

example, the wealth of research on skill or competency-based teacher education. While there exists a reasonable amount of information on problem identification and task-structuring, little is known about the nature of [illegible]ls become integrated into the process of teaching and into the be[illegible]toire as demonstrated in a classroom event. Given such limitatio[illegible]es [illegible]al judgment does at least permit a rich mix of variables to operate [illegible]ging the process of teaching and, at this point, may well be our most valid and reliable measure of effectiveness. At any rate, the data here is encouraging and does suggest a high degree of stability in judgment across years even when one group (students) changed its entire membership each year. Although we can only speculate, the data does suggest some rather consistent definitions of teacher effectiveness were being applied across groups by the evaluators. It would not seem surprising, therefore, if later researchers found that professional judgment as a criterion variable captured some dynamic dimensions of the teaching process which were illusive or left unaddressed in current outcome or student attainment studies. The agenda ahead, and clearly no small one, is for us to learn to trust the criterion variable of professional judgment.

## References

Bentley, R.R. and Price, J. "Purdue Student Teacher Opinionaire." Purdue Measurements and Evaluation Instruments. Purdue Research Foundation, Purdue University, Lafayette, Indiana, 1969.

Hays, William L. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963.

Mahan, James M. and Loadman, William. "A Study of the Relationships Between the Rankings of Supervising Teacher Effectiveness and Attitude Toward Education." Paper presented at the Annual Meeting of AERA, New Orleans, February, 1973.

Siegel, Sidney. Now Parametric Statistics. New York: McGraw-Hill, 1956.

Soar, R.S. and Soar, R.M. Problems in Using Pupil Outcomes for Teacher Education. Washington, D.C.: National Education Association, 1975.

Turner, Richard L. "An Overview of Research in Teacher Education." Teacher Education: The Seventy-Fourth Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1975.