DOCUMENT RESUME

ED 028 129 SP 002 311

By-Leles, Sam; And Others

A Rationale and Related Exploratory Studies for Validating the Ohio State Instructional Preference Scale, Form I.

Pub Date 68

Note-12p.; Paper presented at the annual meeting of the American Educational Research Assn., Los Angeles, California, February 1969.

EDRS Price MF-$0.25 HC-$0.70

Descriptors-Educational Experiments, Education Majors, *Occupational Tests, Predictive Validity, Preservice Education, Screening Tests, *Teaching, *Test Construction, *Test Reliability, *Test Validity

Identifiers-*Ohio State Instructional Preference Scale, OSIPS

The Ohio State Instructional Preference Scale (OSIPS) was designed to identify persons suited or not suited for teaching. After identifying six areas covered in general secondary methods courses, two researchers independently constructed a series of statements representing attitudes, ideas, and dispositions about each area and jointly constructed a 50-item scale and key allowing four possible responses to each item. The instrument was administered on a pre-post basis to beginning education students to measure learning and dispositions before and after a general methods course to determine the direction and extent of changed behavior due to course influences. A series of exploratory studies at Northern Illinois University indicated that OSIPS had potential for predicting achievement in student teaching. Test-retest studies conducted at the University of Alabama (N-122), Northern Illinois University (N-86), and Ohio State University (N-190) supported the reliability of OSIPS. A combination of statistical treatments were performed to assess the validity of individual items and the internal consistency. There is sufficient evidence to justify continued reliance upon the general theory used to construct OSIPS and to support its reliability and validity, but there is an apparent need to modify it by increasing the number of items and the range of responses to each. (JS)

A RATIONALE AND RELATED EXPLORATORY STUDIES FOR VALIDATING THE OHIO STATE INSTRUCTIONAL PREFERENCE SCALE, FORM I

by

SAM LELES, UNIVERSITY OF ALABAMA
C. KENNETH MURRAY, WEST VIRGINIA UNIVERSITY
HARRY R. BARKER, JR., UNIVERSITY OF ALABAMA

COLLEGE OF EDUCATION
ADMINISTRATION, CURRICULUM DEPARTMENT
UNIVERSITY OF ALABAMA
1968

RATIONALE AND RELATED EXPLORATORY STUDIES FOR VALIDATING THE OHIO STATE INSTRUCTIONAL PREFERENCE SCALE, FORM I*

PROBLEM

One of the criticisms of teacher education programs (TEP) is that they encourage persons to enter education on an indiscriminate basis; i.e., screening procedures usually depend upon grade point average and a varying assortment of general education courses. Aside from these considerations, most TEP are set up so that most anyone in the college student body population can enter education programs.

One reason that TEP fail to be selective or to discriminate among those entering programs is a lack of valid and reliable instruments capable of identifying persons suited or not suited for teaching.

This is the problem setting from which the researchers began work to create an instrument. They were interested in assessing two kinds of learnings associated with pre-service teacher candidates. The two kinds of learnings were (1) cognitive information about teaching and learning and (2) attitudes and dispositions which pre-service teachers bring into TEP.

It was reasoned that some attitudes and cognitive notions are congruent with expected learnings in TEP. Others are at odds with these learnings.

People now responsible for TEP make several assumptions about the value of courses in those programs. One assumption holds that the courses offered and taught influence pre-service teacher behavior so that it is evident in subsequent classroom teaching. The second assumption is that course work in TEP is relevant to producing effective classroom teaching when eventually pre-service teachers are certified and teach.

Yet it appears that we are not in a good position to deal scientifically with the second assumption until, first, we possess an accurate picture of the nature and

---

* Paper prepared for and read at the annual meeting of American Educational Research Association in Los Angeles, California, February 5-8, 1969.

impact of existing courses in TEP. In effect, we need a more concise description of what is before we can prescribe what ought to be.

DEVELOPMENT OF THE INSTRUMENT

While at Ohio State University, two of the researchers instructed a general methods course in secondary education. As this was the first course in the professional sequence of the TEP, it was reasoned that any prevailing notions and attitudes about teaching and learning held by pre-service teachers were least contaminated at this point. It was thought that if we could identify the particular areas of learning in such methods courses, then items could be constructed for each of these areas.

It was theorized also that if the instrument was applied on a pre-post basis, one might secure measures of learnings and dispositions which exist prior to a general methods course, and then at the completion of the course. Assuming that the instrument was valid and reliable, this makes it possible to secure measures for areas considered vital to successful teaching before one is exposed to a professional course in education. Securing measures at the close of a course permits examining and determining the appropriateness of learnings thought to be relevant to the methods course. Such measures then can be used to determine the direction and extent of changed behavior due to course influences.

The general secondary methods course at Ohio State University was analyzed so that six areas were identified. The six areas included attitudes, ideas, and dispositions about:

1. the nature of the learner
2. the nature of content
3. the role of teacher as a facilitator of learning
4. measurement and evaluation of learners
5. the objectives of learning
6. the purposes of education

The inclusive nature of these areas suggests that these areas are basic to most general methods courses taught elsewhere. Also, they represent concerns with which pre-service teacher candidates are likely to be acquainted and about which they hold dispositions derived from prior experiences.

Having identified the six components for this general methods course, researchers independently constructed a series or set of statements for each area. Each researcher ranked each statement (or item) for each of the six categories in terms of clarity of meaning and relevance to materials and activities in the methods course. The two researchers then brought together their items and examined each area jointly. By mutual agreement and through a process of elimination, nine items were selected in each of the following areas: "nature of the learner," "nature of content," and "the role of teacher as a facilitator of learning." Eight items were selected for "measurement and evaluation of learners" and for "objectives of learning." With the assignment of seven items to "purposes of education," an instructional preference scale was created with fifty items derived from six areas.

Some sample items in each of the six areas consist of:

Learner

1. The more experiences one has, the broader will be his learning range.
2. Students in courses requiring relatively lower levels of verbalization than that associated with regular secondary classes ought not to be given "A" for achievement at their peak level of capability.

Content

3. In the realm of high school subjects, the subjects that are of greatest value are those most difficult to learn.
4. Content selection is influenced more by cultural forces than by the needs of the learner.

Teacher

5. Teachers should frequently brief students in regard to the students' grade status.
6. What is taught is of considerably greater importance than how it is taught.

Objectives

7. Learning objectives by nature need to be general and long-term.
8. When a student's work is judged as being unacceptable for one reason or another, the teacher should withhold final evaluation until the student has had an additional opportunity(ies) to do it over.

Measurement and evaluation

9. The fairest method of grade distribution is the normal curve.
10. Objective-type tests are to be preferred over subjective types because they eliminate value judgments on the part of the teacher.

Purpose

11. The basic purpose of education is mental growth.
12. Equality of opportunity means that public education must be made available for all persons who are likely to benefit from such education.

The fifty items were arranged sequentially until each of the six areas was repeated at least seven times. Areas with eight and nine items were repeated until the sequence of learner-content-teacher-objectives-measurement-purpose resulted in a fifty-item scale.

The key was constructed by deliberate and arbitrary agreement of the researchers. Each researcher responded to each of the fifty items. Responses were then compared and an r of .60 was calculated. Correlation was calculated by subtracting the number of items for which responses were different (+,-) from the number of items to which responses were in agreement (+,+, or -,-) and dividing the subtrahend by the total number of items. The key was then refined by examining responses provided by Professor John B. Hough who, at that time, was director and coordinator of the general secondary methods program at Ohio State University. Hough's data were calculated for an r of .72. There were four possible responses to each item. Force-choice responses were in terms of agreement-disagreement, and total scores possible for the instrument might range from an H of 250 to an L of 50.

ADMINISTERING THE INSTRUMENT

The purpose of the Ohio State Instructional Preference Scale necessitated defined population samples. Subjects selected for the studies were enrolled in secondary methods courses that were part of a formal teacher education program, TEP. The pre-test was administered during regular class session sometime during the first two weeks of the course, while post-testing occurred during the final week of the quarter or semester.

Subjects were provided the following directions in pre and post-testing:

1. This test and the scores derived from it have no bearing upon student grades for this course.

2. Subjects were asked to record choices on a force-choice basis; that is, indicate slight support or agreement as +1, strong support or agreement as +2, slight opposition or disagreement as -1, and strong opposition or disagreement with -2.
3. The instrument required about 40-50 minutes of time, but subjects were told there was no time limit and they should not work against time to complete it.

FINDINGS

During 1967-68, OSIPS was subjected to a series of exploratory studies at Northern Illinois University. Sample sizes were small, and the purpose of these studies was to explore various treatments of data which might be useful in subsequent efforts to validate the instrument.

A predictive validity study was conducted using secondary student teachers. The sample was divided into two groups exhibiting H achievement (N-12) and L achievement (N-12). Achievement was determined by the student teaching supervisor on the basis of rating sheets. It was hypothesized that the two groups, H and L achieving student teachers, would not differ with respect to mean scores on OSIPS. The mean difference was 20.3 in favor of the H achievement group. The t-value for this difference is 6.72, which is significant beyond the .001 level. On the basis of this data, the hypothesis can be rejected and we might assume that OSIPS has potential for predicting achievement in student teaching when the criterion measure is a rating sheet. However, sample size was small, and studies need to be undertaken to replicate these findings.

A construct validity study was conducted using the Teaching Situation Reaction Test (TSRT). The study was designed to determine whether there was any commonality shared by the TSRT and OSIPS. The TSRT is intended to measure reactions to teaching situations which are intentionally subject matter neutral. The reactions deal with such common aspects of teaching as planning, classroom management, and teacher-pupil relationships. The sample was 36 senior education students in the College of Education at Northern Illinois University. Spearman rank-order correlation coefficient for this sample was .45. The coefficient is significant beyond the .01 level with 35 degrees

of freedom. This modest correlation between the two instruments suggests they do share some commonality.

Three test-retest reliability studies of OSIPS were conducted. The first was computed for a sample, N-37, of pre-service teachers at Northern Illinois University. The test was administered on the first day of class and again six weeks later. The test-retest reliability computed for this sample by product-moment correlation was .82.

Seventy-eight pre-service teachers at Ohio State University participated in the second reliability study. There was an interval of fifty days between pre-post tests. Product-moment correlation of the two sets of scores was .76.

The final test-retest reliability of OSIPS was conducted with data from the University of Alabama sample. Ninety-seven pre-service education students in several sections of general methods courses participated. OSIPS was administered at the start of the course and again six weeks later. The reliability coefficient computed for this sample was .66.

While these three studies provide support for the reliability of OSIPS, it is possible that factors unidentified in these courses also may have affected scores.

Descriptive data on OSIPS are derived from pre-test scores and are calculated for measures of central tendency and dispersion. Data are for three samples of pre-service teachers and are reported in Table 1.

TABLE 1

MEASURES OF CENTRAL TENDENCY AND DISPERSION
FOR THREE SAMPLES ON OSIPS FORM I

SAMPLES	MEAN	MEDIAN	MODE	S.D.	RANGE
University of Alabama, N-122	179	180	182	9.85	51
Northern Illinois University, N-86	180.17	181	176	9.37	45
Ohio State University, N-190	180.85	177	173	11.09	53

At this time, these data have no particular statistical significance. However, the absence of large differences in mean and standard deviation scores suggests support for assumptions stated earlier. The assumptions are that the six areas comprising OSIPS, (1) represent ideas, attitudes, and dispositions familiar to most pre-service teachers, and (2) are basic to most general methods courses.

The next series of studies reported are those conducted at the University of Alabama, 1967-68. Data were secured in 1967-68 from the University of Alabama, N-122, and in 1966-67, from Ohio State University, N-190.

Data were grouped in seven sample combinations, ranging from N-26 to N-312, and factor analyzed using principal component solution with orthogonal varimax rotation. As sampled in six defined areas, findings reveal response patterns (factor loadings) that lack structure. However, response patterns for males are more similar than different as compared to female response patterns. This is suggested by the number of variables with values $>$ .600 in both the Ohio State and University of Alabama samples. Findings based on three samples, N-312, N-126, and N-48, reveal some response patterns as inconsistent with generally accepted principles of learning or research. Principal component factor analysis using orthogonal varimax rotation for seven samples failed to reveal clear structure. Some data suggested several areas holding together, but there is lack of statistical evidence for identifying specific items in OSIPS.

Post-test data for male and female were treated separately to factor analysis. However, separate treatment for male and female data resulted in an unanticipated difficulty. Samples were reduced excessively due to lack of post-test scores for many subjects. Factor analysis of post-test scores for which there were corresponding pre-test scores, was based on N-174. This was a reduction of 138 from the 312 pre-test scores.

It was hypothesized that the factor matrix derived from these data, N-174, would reveal six variables that corresponded to the six sub-test areas used to construct OSIPS. The factor matrix for pre-test scores extracted two factors. Clear

structure was not evident in either treatment and the hypothesis was rejected.

Although the hypothesis was rejected, there was appearance of several sub-test areas holding together. Using post-test data, researchers examined the internal validity and reliability of OSIPS. To assess the validity of individual items, performance on each item was correlated (Pearson Product Moment) with total score on the sub-test to which the item belonged. Correlations were low to moderate in size but, excepting one item, were statistically significant at the .05 level. This provides evidence that test items are measuring what subtests as a whole are measuring.

The alpha coefficient of internal consistency (Cronbach's KR-20) was applied to OSIPS to assess the reliability among items of each of the six subtests. Reliabilities were uniformly low, owing to the small number of items comprising each subtest. It may be noted that the KR-20 formula provides an indication of the average correlation of all possible correlations based on different split halves of each subtest.

OSIPS was then studied by the Cattell Pattern Similarity Index (Rp).[1] Pre-post test data were compared for N-174, and raw mean scores for each subtest area were computed. Profiles for each of the subtest Z-score means were compared. Table 2 shows the profile mean scores with corresponding (pre-post) standard deviations.

TABLE 2

PRE AND POST RAW SCORE MEANS OF SIX PROFILE ELEMENTS IN OSIPS-FORM I, N-174

GROUP	MEAN PROFILE SCORES					
	1 LEARNER	2 CONTENT	3 TEACHER	4 OBJECTIVES	5 MEASUREMENT	6 PURPOSE
PRE-TEST	36.16	28.21	37.75	30.15	26.94	21.06
S	3.39	3.98	3.21	3.80	4.02	3.02
POST-TEST	36.40	27.55	37.23	28.98	26.02	21.26
S	2.88	3.80	3.38	3.67	3.80	3.23

[1] R. B. Cattell, "Rp and Other Coefficients of Pattern Similarity," PSYCHOMETRIKA, vol. 14:no. 4, 1949, pp. 279-298.

Differences in pattern between pre-test profile means and post-test profile means was computed for a coefficient (rp) of -0.557 which is statistically significant beyond the 0.01 level. This coefficient indicates that the pattern of subtest means of OSIPS changed significantly between the pre and post tests.

Treatment of profiles of subtest data by the Cattell Pattern Similarity Index and determination of item and subtest consistency by product-moment correlation contributed support for the internal validity of OSIPS. However, the subjective nature of each treatment was recognized and acknowledged.

Finally, a more objective examination of profile data was undertaken by use of Ward's Hierarchical Grouping Technique.[2] As stated earlier, the construction of OSIPS was predicated on the notion that each of the six subtest groupings is homogeneously composed. As an alternate method useful to assessing the internal validity of OSIPS, the differences in profiles between response patterns of all subjects for items one and two are compared, first; then items one and three; then items one and four and so on. Items with identical or most alike response profiles are eventually grouped together to form homogeneous assortments of items. Post-test data treated by Ward's Hierarchical Grouping Technique identified correctly the majority of items with appropriate subtests in five areas. Grouping for the sixth area remained undefined as neither "content" items nor "objective" items established a majority. Treatment of pre-test data by the same technique identified only several of the subtest areas and showed considerable overlap of items. At this stage of its development, the internal validity of OSIPS is provided additional support in treatment of data by Ward's Hierarchical Grouping Technique.

---

[2]J. H. Ward, Jr., "Hierarchical Grouping to Optimize an Objective Function," JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, vol. 58, 1963, pp. 236-44.

DISCUSSION

The overall and cumulative effect of these studies suggests sufficient evidence to justify continued reliance upon the general theory used to construct OSIPS. However, these studies do not provide the empirical evidence needed to identify particular items that support the internal validity of OSIPS. Through a combination of statistical treatments, there is evidence and support for the reliability and validity of the instrument.

Difficulty in identifying structure in OSIPS may be due, in part, to the limited number of items used to construct each subtest area of the instrument. This suggests the possible need to increase the number of items for each subtest area in OSIPS. Also, low correlations for the subtests may be due to the relatively limited number of responses afforded respondents for each item. There is need to consider enlarging the range of item responses.

These studies show the complexity in devising and attaining valid instrumentation directed to identifying subjects suited or not suited for TEP. The problem is especially difficult when an attempt is made, as it was in this case, to deal simultaneously with "knowledge, attitudes, and dispositions." The results of these studies are sufficient, however, to warrant continued research and experimentation with OSIPS. While the theory supporting the instrument appears adequate, there is apparent need to modify OSIPS in terms of:

1. increasing the range of responses for each item,
2. enlarging the instrument as a whole by increasing the number of items for each of the six subtest areas composing OSIPS, and
3. revising, omitting, and adding items as needed.

REFERENCES

Cattell, R. B. "Rp and Other Coefficients of Pattern Similarity," Psychometrika. Vol. 14:No. 4, 1949, pp. 279-298.

Ward, J. H., Jr. "Hierarchical Grouping to Optimize an Objective Function," Journal of the American Statistical Association, Vol. 58, 1963, pp. 236-244.