DOCUMENT RESUME

ED 317 553 TM 014 567

AUTHOR Carifio, James; Kermis, William
TITLE Some Findings on Preparing Teachers' Attitudes towards Tests.
PUB DATE Feb 90
NOTE 38p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Tampa, FL, February 13-17, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Attitude Measures; *Education Majors; Evaluation; Higher Education; Mathematics; *Teacher Attitudes; *Test Use; *Undergraduate Students
IDENTIFIERS *Teacher Candidates; Teacher Surveys

ABSTRACT

Six studies concerning the attitudes of preparing teachers toward tests, evaluation, and mathematics are reported. These studies report scale development, validations, cross-validations, and observed results. Three instructors participated in the development of an attitude survey in the first study; a total of 923 undergraduate preparing teachers participated in the five other studies (291, 226, 54, 136, and 156, respectively). Preparing teachers had more positive attitudes toward evaluation and mathematics than toward tests. No correlations were found between their attitudes toward evaluation, which were positive, and their attitudes toward tests (generally negative) and mathematics (slightly positive), but there were moderately significant correlations between their attitudes toward tests and mathematics. High-aptitude preparing teachers had significantly more positive attitudes toward mathematics and tests than did low-aptitude preparing teachers. The locus of the observed differences was primarily on affective dimensions of these attitudes. Implications are discussed in terms of theory, education, and instructional practice, and in terms of a study of the approach of instructors by attitude. Ten tables present data from the studies. (Author/SLD)

SOME FINDINGS ON PREPARING TEACHERS' ATTITUDES TOWARDS TESTS

James Carifio, University of Lowell
William Kermis, Northwest Missouri State Univeristy

Abstract

This paper reports 6 studies on preparing teachers' attitudes towards tests, evaluation and mathematics. The six studies report scale development, validations, cross-validations, and observed results.

Preparing teachers were found to have more positive attitudes towards evaluation and mathematics (in this order) than towards tests at the .001 level of significance. No correlations were found between preparing teachers' attitudes toward evaluation, which were positive, and their attitudes towards tests and mathematics, but moderately significant correlations were found between preparing teachers' attitudes towards tests and their attitudes towards mathematics.

High aptitude preparing teachers had significantly more positive attitudes towards mathematics and tests (in this order) that low aptitude preparing teachers, but the attitude levels of both high and low aptitude preparing teachers towards tests were in general negative, whereas for mathematics it was slightly positive. The locus of the observed differences observed were primarily on the affective dimensions of these two attitudes.

Tests are a very potent affective stimulus for preparing teachers. The implications of these results are discussed in terms of both theory, education, and current instructional practices in teacher education, and in terms of a seventh study conducted and reported in this paper.

Paper presented at the annual meeetung of the Eastern Educational Reasearch Association, Tampa, Florida, Feb. 13-17, 1990

Although lively discussions and subjective reports about teachers' and preparing teachers' attitudes towards tests go back to before the turn of the century (see Kirland, 1971), little empirical research has been done on this topic other than a few seminal and very limited studies which collected unstructured interview data or responses to 5 or 6 Likert type opinion items about tests and the impact their uses in education (e.g., Osborne, 1933; Tyler, 1936; Noll, 1956; Mayo, 1967; Goslin, 1967; and Marsullio, 1971). As Evans (1983) has said, for a topic that that is stressed so strongly by so many, and seems to concern professional educators so greatly, very little is actually known empirically, other than the general belief that teachers' and preparing teachers' attitudes towards tests and testing are in general very negative (see Mayo, 1967; Goslin, 1967; and Kirland, 1971).

Statement of the Problem

One of the major reasons that so little is known empirically about teachers' and preparing teachers attitudes towards tests is the lack of appropriate instrumentation and flexible procedures for measuring these attitudes that can be used in a broad range of different research situations. The purpose of the present paper, therefore, is to report the development of a semantic differential (Osgood et al., 1964) for measuring teachers' attitudes towards tests, and some results that were obtained in several studies of preparing teachers' attitudes towards tests conducted with this instrument and other measures.

Semantic Differentials and Attitudes

The range of evidence supporting the semantic differential technique as a measure of attitudes is considerable (Nunnally, 1967). Scores from semantic differentials tend to correlate highly with scores from equivalent-form Likert scales which is strong evidence of their validity (Cronbach, 1970). However, unlike Likert Scales and similar attitude measuring procedures, semantic differentials have the advantage of being quick and easy to administer in a wide variety of situations, and the "items" on the scale tend not to become "dated" which is another important characteristic. Further, semantic differentials also have another major and outstanding advantage over other attitude measuring techniques that were not well-perceived in the sixties and seventies in this country; namely, if constructed correctly, semantic differentials are analog measures of a person's cognitive and affective schema of the psychological object in question, which allows data from semantic differential scales to directly and easily related to cognitive and information processing theories of learning and change. This point is somewhat obvious to modern cognitive and information processing learning theorists, but it will be explained in more detail below.

Attitudes are generally defined as a person's feelings, opinions, views, and judgements about particular entities, concepts, or psychological objects (Guilford, 1965; McGuire, 1968; Cronbach, 1970). The semantic differential technique presents a person with a general concept word (such as tests) and a series of bipolar adjective-pairs (such as valuable-worthless) placed at each end of a seven category anchored continuum for the person to express her or

his attitude by marking the appropriate category on the continuum.

Typically, two major underlying dimension are found when semantic differential responses are factor analyzed to assess their construct validity (Guilford, 1965; Nunnally, 1967). These two dimensions are called the evaluation (or cognitive) dimension and the potency (or affective) dimension. These two basic dimensions reflect the two basic constructs or factors identified in most theories of attitude found in the literature (e.g., Peak, 1955; McGuire, 1967; Anderson, 1970).

Usually, the two underlying factors found on semantic differentials are scored by a simple summation procedure to obtain two subscale scores and then a total attitude score, each of which is then divided by the number of items that enter into each summation to put the scores on a scale that is easily interpretable and of the same metric. Several theorists have recommended using factor scores (e.g., Harris, 1965), and others have recommended using only total scores (e.g., Guildford, 1965). Both of these recommendations, however, have a number of problems associated with them when the assessment of attitude change is the focus and more than one attitude is examined (Cronbach, 1970).

In a comparative study of these issues, Richards (1972) found that logical scoring of semantic differentials and use of simple pre-to-post difference scores introduced no bias or distortion into results as compared to all of the other suggested procedures, and had the advantages of being both simple to do and easy to interpret. Given Richards' results, all of the semantic differentials used in the present studies were scored by the simple procedures outlined

above.

Study 1

The semantic differential development procedures we used were different from those recommended by Osgood (1964). We followed the procedures recommended by Nunnally (1967) and Cronbach (1970) which seek to maximize the content and construct validity of the resulting scale relative to traditional theories of attitudes. The attitude theory we used was classical attitude theory (Peak, 1955; McGuire, 1968). As stated above, this theory predicates two underlying dimensions for attitude which are the cognitive (or evaluative) dimension and the affective (or potency) dimension. The cognitive dimension in this view describes beliefs, opinions, and judgements, while the affective dimension describes emotions, feelings, drives and approach-avoidance tendencies. This model of attitudes guided all of the activities, studies and analyses that are reported below.

Three highly experienced test and measurements instructors (none of whom were the authors of this paper) in a large university in the northeast selected and ordered 15 cognitive (evaluative) and 15 affective (potency) adjective-pairs from a pool of 75 adjective pairs that all three agreed reflected the range and variation of feelings, opinions, views, and judgements that teachers, preparing teachers, and people in general tend to express about tests. The sources of the adjective pairs in the initial pool were (1) Osgood's (Snider and Osgood, 1969) list of validated pairs, (2) free associations (see Garskoff and Houston 1963; Johnson, 1971; and Milligan, 1983) made to the stimulus word tests by teachers, preparing teachers and others

(N=74) rank ordered by frequency of occurrence, and (3) adjectives (and their implicit opposites) that tended to occur frequently in a sampling of the extant "pro and con" literature on tests and testing.

What is important about this study is that there was a high degree of cofrequencies on one or both adjectives in a final 30 pairs "piloting" list derived between the three adjective sources, with 20 of the 30 pairs selected being on all three lists. The remaining 10 adjective pairs selected adjectives that occured very frequently and tended to come up first in preparing teachers' free associations about tests. These 10 adjective pairs also were on the "pro and con" list from the literature.

For a very great many preparing teachers, affective (highly emotional) adjectives tended to come up first, at a rate of roughly 3 to 1, as compared to cognitive adjectives, during the first two minutes of free association, whereas for a great many other preparing teachers' cognitive adjectives came up first, in at a rate of roughly 4 to 1, as compared to affective adjectives, during the first 2 minutes of free association. This finding is very important for a great number of reasons, but this finding cannot be elaborated upon more fully until the end of this paper, as other data are needed to understand and explain the full significance and importance of this finding.

The unselected 45 adjective pairs from this first study represent features (see Neiser, 1967 and Norman, 1981) of tests, testing, and attitudes towards tests that are less important phenomonologically than the 30 adjective pairs that were retained. This fact in no way necessarily implies that any one of the rejected

pairs are not important features or dimensions of tests, testing, or attitudes towards tests. It simply means that these adjective pairs were members of a latent features category rather than the dominant features category. Features can, of course, shift categories with time which in and of itself would provide useful and interesting research data on this problem (see House, 1954, Tulving, 1962, and Shavelson, 1971 for details) that can be generated in the future by comparing free association and other types of results to the adjective pairs on the final version of the semantic differential developed and reported on in detail below.

The reasons that we have not reported the results of this first study in any more detail than is reported above are (1) it was a pilot/developmental study that was not designed to be a true experiment but rather a quasi-experiment to formulate the universe and problem domain, (2) the results of this study cannot be neatly, directly, and tightly tied to the final semantic differentials that were arrived at for use in our other studies, and (3) detailed reporting of this study is simply not germane to the main goals and purpose of this paper. The key points of this first study, however, are (1) the underlying Critical Features and cognitive theories of attention, arousal, perception, spreading activation, and subsequent information processing (see Norman, 1981 and Milligan, 1983) that we used in instrument development along with classical theories, (2) the process we used to arrive at 30 adjective pairs to investigate empirically, and (3) what we initially found as a result of this process.

Study 2

Data were collected from 291 volunteer undergraduate preparing teachers at a large university in the northeast on the 30 adjective-pairs selected by the process described in study 1 above, prior to their taking an introductory course in educational tests and measurements. As part of this data collection procedure, data were also collected from these 291 subjects on a semantic differential which had as its stem the phase "To me, evaluation is" which was followed by the exact same 30 adjective pairs as the semantic differential that had as its stem the phase "To me, tests are." A semantic differential that measured attitudes towards mathematics developed by McCallon and Brown (1971) was also part of this data collection package. The McCallon semanatic differential had 15 adjective pairs, of which only 2 were the same as the 30 adjective-pairs for the semantic differentials on attitudes towards tests and evaluation. The order in which these semantic differentials were administered to subjects in this study were attitude towards tests, evaluation and then mathematics.

Using standard principle component analysis, McCallon and Brown (1971) found a two factor (evaluation and potency) structure for their 15 item semantic differential for 68 undergraduates that accounted for 67% of the total variance. Scores from their semantic differential correlated with scores from the Aiken and Dregar (1965) attitude towards mathematic scale at r=+.90. The latter scale was a 40 item validated Likert scale.

The reason that this experimental design was used in this study

was to implement beginning at this point a Campbell and Fiske (1967) convergent and discriminant validation design. The McCallon and Brown instrument and their findings was the "known marker," the semantic differential and attitude towards evaluation was the "control stimulus," and the attitudes towards test semantic differential was given first to keep the responses subjects made to it uncontaminated in any way. The other reason that this design was used is that we were trying to conceptualize, formalize, and implement a long term and on-going research program in this general area, and all of these instruments were needed for such a program. We personally would have preferred a scale that measured attitude towards statistics rather than (or as well as) mathematics, but there is currently no such scale in existence, and a "known marker" was needed to produce strongly confirmatory results in these studies.

The data collected in this study were factor anlayzed using principle component analysis with unities in the diagonals, an eigen cut-off value of 1.0 and varimax rotation (Harmond, 1965; Tatsuoka, 1988). Three factors were found for attitudes towards tests using this "classic and conservative" factor analysis model that accounted for 83% of the total variance. The first factor was the cognitive factor (44%) and the second was the affective factor (28%). The third factor was made up of those adjective-pairs that had very little variation in student responses. Virtually the same results were found for the attitudes towards evaluation scale with the same adjective pairs constituting the same three factors and accounting for 77% of the variance. Two factors were found for the McCallon scale which accounted for 68% of the total variance and these two

factors were the same as the factors found by McCallon and Brown (1971). When the adjective-pairs of all three semantic differentials were factored together as one unit, eight factors were found that were the same as the eight factors reported above. The meaning and significance of the findings will be commented on more fully in study 3 below.

The 10 adjective-pairs that loaded most strongly (>.60) on the cognitive component and the 7 pairs that loaded most strongly (>.60) on the affective component for attitude towards tests and evaluation were retained for further study. These 17 adjective-pairs were the same for attitude towards tests and evaluation. Retained pairs were left in the order they had been placed in on the original 30 pair list.

Two further precautions were taken at this point. First the frequency of neutral response to the 17 retained adjective-pairs were analyzed to make sure that the results found were not due to a preponderance of neutral responses to these items. All 17 adjective-pairs retained had very low neutral responses, the highest being 2.6% of the 291 responses made.

The second precaution taken is that two "constant response" adjective pairs were retained on the final scales as an internal validity check for any administration of the scales and to break response set in responding to the scales. These two "constant" adjective-pairs were "risky-safe" and "secure-dangerous (reverse scored)," both of which had factor loading greater than .90 on the third (constant) factor found. The first of these two "constant" adjective pairs were retained as item 1 on the scales and the second

was retained as item 17 on the scales. Therefore, the attitude towards tests and evaluation semantic differentials administered in the studies reported below had 19 items on them, 17 of which were used to generate score and 2 of which were used to internally check the quality of scores. The reason why the factor structures found in this study and the list of adjective-pairs retained are not reported here is that they will be reported in the next study which is the confirmatory study of the scales.

Study 3

In the next study, data were collected from 226 volunteer undergraduate preparing teachers at a large university in the northeast on the 17 retained adjective-pairs for attitudes towards tests and evaluation and McCallon and Brown's (1971) attitude towards test scale, prior to these preparing teachers taking an introductory course in educational tests and measurements. The order in which these three semantic differentials were administered were attitudes towards tests, evaluation and then mathematics for all students.

The data collected in this study were factor analyzed using principle component analysis with unities in the diagonals, an eigen cut-off value of 1.0 and varimax rotation (Harmond, 1965; Tatsuoka, 1988). As can be seen from Table 1, a two factor structure for attitude towards tests was obtained for 17 "scored" adjective-pairs which accounted for 66% of the variance. The two factors obtained were the evaluation (or cognitive) factor which accounted for 40% of the total variance and the potency (or affective) factor which

Table 1: FACTOR STRUCTURE OF THE ATTITUDE TOWARDS TESTS SEMANTIC DIFFERENTIAL (N=226)

ADJECTIVE PAIR	I	II	h2
meaningful: meaningless	.79	.34	.73
concealing: revealing	.74	.28	.63
necessary: unnecessary	.81	.18	.69
useful: useless	.82	.34	.79
oppressive: liberating	.38	.65	.56
effective: ineffective	.78	.36	.74
boring: interesting	.28	.68	.53
beneficial: harmful	.70	.38	.63
frustrating: stimulating	.23	.77	.63
valuable: worthless	.81	.29	.75
pleasant: unpleasant	.25	.78	.67
sinister: intriguing	.25	.75	.50
important: unimportant	.81	.29	.74
repugnant: likable	.24	.80	.70
needed: unneeded	.75	.33	.67
helpful: unhelpful	.72	.40	.68
satisfying: unsatisfying	.48	.57	.55
N = 226 Variance	40%	26%	66%

I = Cognitive Dimension II = Affective Dimension

Table 2: FACTOR STRUCTURE OF THE ATTITUDE TOWARDS EVALUATION SEMANTIC DIFFERENTIAL (N=226)

ADJECTIVE PAIR	I	II	h2
meaningful: meaningless	.77	.23	.65
concealing: revealing	.33	.45	.31
necessary: unnecessary	.80	.18	.68
useful: useless	.81	.18	.68
oppressive: liberating	.38	.41	.34
effective: ineffective	.52	.41	.45
boring: interesting	.14	.74	.56
beneficial: harmful	.75	.17	.59
frustrating: stimulating	.19	.79	.65
valuable: worthless	.84	.22	.75
pleasant: unpleasant	.16	.74	.58
sinister: intriguing	.08	.52	.27
important: unimportant	.83	.22	.74
repugnant: likable	.18	.75	.59
needed: unneeded	.83	.19	.72
helpful: unhelpful	.83	.19	.62
satisfying: unsatisfying	.42	.47	.39
N = 226 Variance	35%	22%	57%

I = Cognitive Dimension II = Affective Dimension

accounted for 26% of the total variance. These two factors and the adjective-pairs which comprise each of them are very clear and straightforward as can be seen from Table 1. These two factors and the adjective pairs that comprised each of them, moreover, conformed with the expectations and framework that was established in study 1 without exception.

All 17 adjective-pairs on the attitudes towards tests semantic differential predicted total score at the .001 level of significance with the item-total correlations ranging from .48 and .83. The mean item-total correlation was .61 with a standard deviation of .17. The median item-total correlation was .64. The Cronbach's alpha internal consistency coefficient was r=+.87 for the 17 items.

Table 2 presents the factor structure for the attitudes towards evaluation. As can be seen from Table 2, a two factor structure for attitude towards evaluation was obtained for 17 "scored" adjective-pairs which accounted for 57% of the variance. The two factors obtained were the evaluation (or cognitive) factor which accounted for 35% of the total variance and the potency (or affective) factor which accounted for 22% of the total variance. These two factors and the adjective-pairs which comprit each of them are not as c'ear or straightforward as those for attitudes towa-ds tests which are given in Table 1. These two factors for attitudes towards evaluation and the adjective pairs that comprised each of the two factors, however, do conform with the expectations and framework that was established in study 1 with four exception.

All 17 adjective-pairs on the attitudes towards evaluation semantic differential predicted total score at the .01 level of

significance with the item-total correlations ranging from .21 and .76. The mean item-total correlation was .51 with a standard deviation of .23. The median item-total correlation was .53. The Cronbach's alpha internal consistency coefficient was r=+.78 for the 17 items.

Table 3 presents the factor structure for attitudes towards mathematics. As can be seen from Table 3, a two factor structure was obtained for the 15 adjective-pairs which accounted for 70% of the variance. Again, the two factors obtained were the evaluation (or cognitive) factor which accounted for 43% of the total variance and the potency (or affective) factor which accounted for 27% of the total variance. These two factors for attitudes towards mathematics and the adjective pairs that comprised each of the two factors, conformed approximately to the structure found by McCallon and Brown (1971). The differences found are due most probably to our much larger sample size.

All 15 adjective-pairs on the attitude towards mathematics semantic differential predicted total score at the .001 level of significance with the item-total correlations ranging from .39 and .72. The mean item-total correlation was .52 with a standard deviation of .21. The median item-total correlation was .54. The Cronbach alpha internal consistency coefficient was r=+.88 for the 15 items.

When the adjective-pairs of all 3 semantic differentials were factor analyzed together as a unit, six factors were obtained, just like the results that occurred in study 2, which corresponded to the underlying dimensions of the 3 scales given in Tables 1, 2, and 3.

Table 3: FACTOR STRUCTURE OF THE ATTITUDE TOWARDS MATHEMATICS SEMANTIC DIFFERENTIAL (N=226)

ADJECTIVE PAIR	I	II	h2
pleasant: unpleasant	.53	.72	.80
bad:good	.85	.27	.81
hard: soft	.14	.79	.63
afraid: unafraid	.40	.69	.67
active: passive	.81	.15	.68
valuable: worthless	.75	.12	.57
strong: weak	.85	.10	.73
love:hate	.32	.81	.76
fast:slow	.62	.17	.41
comfortable: uncomfortable	.49	.71	.74
awful: nice	.75	.47	.78
enjoyable: unenjoyable	.73	.50	.78
light: heavy	.82	.07	.67
varied: repetitive	.56	.41	.45
secure: insecure	.54	.74	.83
N = 226 Variance	43%	27%	70%

I = Cognitive Dimension II = Affective Dimension

Table 4: INTERCORRELATIONS BETWEEN ATTITUDES TOWARDS TESTS, EVALUATION, AND MATHEMATICS (N=266)

	F1T	F2T	TotT	F1E	F2E	TotE	F1M	F2M	TotM
F1T	1.0	.58*	.91*	.08	.02	.03	.31*	.10	.23*
F2T		1.0	.85*	.05	.05	.04	.12	.38*	.30*
TotT			1.0	.05	.02	.04	.25*	.23*	.27*
F1E				1.0	.54*	.87*	.09	.03	.08
F2E					1.0	.81*	.02	.11	.07
TotE						1.0	.01	.11	.09
F1M							1.0	.64*	.94*
F2M								1.0	.90*
TotM									1.0

*P<.01

KEY:
F1T=Factor 1 Attitudes towards Tests
F2T=Factor 2 Attitudes towards Tests
TotT=Total scale Attitude towards Tests
F1E=Factor 1 Attitudes towards Evaluation
F2E=Factor 2 Attitudes towards Evaluation
TotE=Total scale Attitude towards Evaluation
F1M=Factor 1 Attitude towards Math
F2M=Factor 2 Attitude towards Math
TotM=Total scale Attitude towards Math

These factors account for 66% of the variance. A Skree plot of the eigen-values and other statistical considerations, however, indicated that a cut-off value of 1.63 was a truer cut-off value that the "classic and conservative" cutt-off value of 1.0. The entire set of adjective-pairs, therefore, were refactored by principal components analsyis with unities in the diagonals and an eigen cut-off value of 1.63. When this was done, five factors were found that account for 62% of the variance; namely, the McCallon attitude towards mathematics scale collapsed into one factor. A variety of oblique factor analyses of individual semantic differentials and all three semantic differentials together indicated that the factors found were moderately to slightly correlated, but not statistically reducible with the exception of the McCallon scale in the larger context.

These findings are not atypical of attitudinal data or factor analyses of such data. Factor analysis is, in the main, a construct validity procedure (Kerlinger, 1983) that gives static statistic (internal) structure at a given time point. If principal component orthogonal factors scores were generated for the factors of these three semantic differentials, they would be uncorrelated which means that these three attitudes and their underlying factors are uncorrelated and independent of each other, which is not only extremely strong construct validity, but also an important experimental finding in and of itself, which cannot be explained more fully until study five because other data is needed to understanding this finding in detail. However, because factor analysis gives a static internal and relative structure at a given time point, factor scores are both useless and fraught with innumerable difficulties and

problems across time points, and in determining mean levels of responses on any given factor at any given time point, as the mean of any factor score is always zero as the score in a regressed internalized z-score for the factor. It is for this very reason, therefore, that factor scores are not used to develop scores for the "factors" of a given scale, and that the "simple raw summation of the items that comprise the factor divided by the total number of items for the factor" procedure is used that was outlined above (Richards, 1972). This procedure allows directly comparable mean levels on factors to be observed, but the factors will not necessarily be uncorrelated. Any correlations that are observed between factors when "simple summative" scoring procedures are used to generate factor scores, are typically called then "natural or real world" correlations that are in the data that describe the underlying dynamics that are in operations between the factors (Richards, 1972).

Table 4 presents the intercorrelations that were observed between attitudes towards tests, evaluation and mathematics in this study. The factor scores and total scores used to compute the correlations given in Table 4 were generated by the simple summation technique described above. As can be seen from Table 4, the intercorrelations ranged from $r=+.02$ ($p>.05$) between subscales scores to $r=+.27$ ($p<.001$) for total scale scores. What is very clear in Table 4 is that preparing teachers' attitudes towards tests are completely uncorrelated to their attitudes towards evaluation, but slightly correlated to their attitudes towards mathematics. Preparing teachers' attitudes towards evaluation were also uncorrelated to their attitude towards mathematics. These findings are indeed rather

startling and extraordinary, but once again a full explanation and interpetation of these findings is postponed until study 6, as other very important data are needed to understand and clearly explain the full significance and importance of these finding.

It should also be noted that this third study was primarily a confirmatory and scale norming study. Consequently, given all of the results presented above, it would seem warranted to conclude that the final forms of the attitudes towards tests, evaluation and mathematics semantic differentials used in this study had a high degree of construct validity relative to the attitudes that they were designed to measure.

Study 4

A test-retest reliability study was done on the final forms of the attitude towards tests, evaluation, and mathematics semantic differentials described in study 3 above. Subjects were 54 volunteer undergraduate preparing teachers at a large university in the northeast who had not yet had a course in educational tests and measurements. The time interval between the test and the retest was 14 days.

Table 5 presents the test-retest reliability coefficients obtained in this study. As can be seen from Table 5, the test-retest reliability coefficients ranged from r=+.86 to r=+.95. Total scale test-retest reliabilities were r=+.90 for attitude towards tests, r=+.88 for attitude towards evaluation, and r=+.94 for attitude towards mathematics. All three semantic differentials had excellent test retest reliability coefficients for 15 to 17 item scales.

Table 5: TEST-RETEST RELIABILITY COEFFICIENTS FOR ATTITUDES TOWARDS TESTS AND ATTITUDES TOWARDS MATHEMATICS (N=54)

	K	N	Rtt
Factor 1, Tests	10	54	.87
Factor 2, Tests	7	54	.91
Total Scale, Tests	17	54	.90
Factor 1, Evaluation	10	54	.86
Factor 2, Evaluation	7	54	.89
Total Scale, Evaluation	17	54	.89
Factor 1, Math	9	54	.95
Factor 2, Math	6	54	.92
Total scale, Math	15	54	.94

K=number of items

Table 6: CONVERGENT AND DISCRIMINANT VALIDITY COEFFICIENTS BETWEEN THE SEMANTIC DIFFERENTIALS AND THE LIKERT TYPE VERSIONS OF EIGHT OF THE ATTITUDES TOWARDS TESTS ADJECTIVE-PAIRS (N=196)

	Likert Form (8 items)		
Semantic Differentials	F1T	F2T	TotT
Factor 1, Tests	.78	.64	.76
Factor 2, Tests	.79	.74	.65
Total Tests	.73	.74	.74
Factor 1, Evaluation	.28	.24	.24
Factor 2, Evaluation	.13	.19	.19
Total Evaluation	.23	.24	.24
Factor 1, Mathematics	.33	.31	.33
Factor 2, Mathematics	.26	.38	.37
Total Mathematics	.30	.31	.35

All correlations are significant at the .05 level

Study 5

To determine the semantic validity and meaningfulness of the 17 adjective-pairs on the attitude towards tests scale, 8 adjective pairs (4 cognitive and 4 affective chosen at random) were converted to Likert type attitude statements that were included in random order on a 30 item end of course questionnaire for an introductory course in educational tests and measurement. Examples of the Likert type statements used were, "Tests conceal more about people than they reveal," "Tests are an unimportant part of education," and "tests have an oppressive effect on people." Further, it should be noted that in addition to assessing the validty and meaningfulness of the adjective-pairs on the attitude towards tests semantic differential, we also wanted to have an estimate of how versatile these adjective pairs were across different item formats for future research in this general area.

Subjects in this fifth experiment were 196 undergraduate preparing teachers at a large university in the northeast. These preparing teachers were given the three semantic differentials in their final form in the standard order described above at the end of their second to last class. At the end of their last class a week later, these preparing teachers were given the university's standard course evaluation form and then the 30 item course questionnaire described above.

Table 6 presents the intercorrelations observed between the three semantic differentials and the Likert type versions of 8 of the attitude towards tests adjective-pairs (4 cognitive and 4 affective

chosen at random). As can be seen from Table 6, the same patterns and levels of correlation coefficients were observed between the three semantic differentials and the Likert type versions of the 8 attitude towards tests adjective-pairs as were observed and presented in Table 4, between the three semantic differentials themselves. Given this point, it would seem warranted to say from this second convergent and discriminant validity study that preparing teachers semantically interpret the adjective-pairs on these three semantic differential scales in the manner and with the meanings that the adjective-pairs themselves imply.

When the 30 item Likert type questionnaire was factor analyzed using principal component analysis with unities in the diagonals and eigen cut-off value of 1.0, the 4 Likert cognitive items for attitudes towards tests and the 4 Likert cognitive items for attitude towards tests formed two separate factors, with the other items of the questionnaire forming their own factors. In all, four clearly interpretable factors that accounted for 66% of the variance were found. Preparing teachers semantically interpret the adjective-pairs on these three semantic differential scales in the manner and with the meanings that the adjective-pairs themselves imply.

Study 6

In this study, 156 undergraduate preparing teachers at a large university in the northeast were given the 3 semantic differentials described above in the standard order described above. Combined 12-th grade SAT scores were obtained from admissions records for these students. The median combined SAT score for this group (998)

was used to divide these preparing teachers into "high" and "low aptitude" groups (see Table 7 for all relevant statistics).

Multivariate analyses of variances (MANOVAS) were performed on the simple sum subscale and total scale fscores for these three attitudes by aptitude level using Jerome Finn's Univeristy of Buffalo MANOVA programs. As can be seen from Table 8, significant difference at the .001 level were found between high aptitude and low aptitude preparing teachers on the affective dimension of attitudes towards tests and on the cognitive and affective dimensions of attitude towards mathematics. No differences were found between high and low aptitude preparing teachers on attitude twoards evaluation or on the cognitive dimension of attitude towards tests. The significant differences between high and low aptitude students on the affective dimension of attitude towards teste "carried over" to total scale scale as can be seen in Table 8. The omega-squares (explained variances) for these F-ratios were 24.7% and 27.3% respectively which means that aptitude level accounted for a very significant amount of the variance observed in preparing teachers' attitudes. These results may be understood more fully from Tables 9 and 10.

As can be seen from the means and standard deviations given in Table 9, preparing teachers were very positive in their attitudes towards evaluation, with a mean response of 5.1 on a 7 point scale. However, as a group, preparing teachers were significantly more positive in their attitudes towards mathematics ($p<.01$) than they were in their attitudes towards tests. Their mean attitude towards math was 4.3, whereas their mean attitude towards tests was 3.6.

As can be seen from Table 10, preparing teachers' attitudes

Table 7: MEAN SCHOLASTIC APTITUDE TEST SCORES OF SUBJECTS AND HIGH AND LOW APTITUDE GROUPS (N=156)

	n	Mean	St. Dev.
Math	156	499.4	86.7
Verbal	156	494.8	82.8
Total	156	991.1	154.5
High Aptitude	78	1,113.7	84.4
Low Aptitude	78	874.3	89.8

median for median split: 998.0

Table 8: F-RATIO FOR MANOVAS ON THE SEMANTIC DIFFERENTIAL ATTITUDE SCORES BY HIGH AND LOW APTITUDE LEVELS (N=156)

Variable	Aptitude Level
F1T: CD ATT	1.9
F2T: AD ATT	7.2*
F1E: CD ATE	0.1
F2E: AD ATE	0.1
F1M: CD ATM	37.6***
F1M: AD ATM	23.2***
Mult.-F (df=5,140)	10.4***
omega square	24.7%
TotT: TATT	5.1*
TotE: TATE	0.1
TotM: TATM	64.1***
Mult.-F (df=2,140)	12.6***
omega square	27.3%

*=<.05 **=<.01 ***=<.001

Table 9: ATTITUDE MEANS AND STANDARD DEVIATIONS FOR HIGH AND LOW APTITUDE LEVELS (N=156)

Variable:	N	Attitudes Towards Tests: CDATT Mean	CDATT SD	ADATT Mean	ADATT SD	Total ATT Mean	Total ATT SD
High Apt	78	4.2*	1.0	3.3	0.7	3.8	0.8
Low Apt	78	4.1	1.1	3.0	0.8	3.4	0.9
Total	156	4.1	1.1	3.1	0.8	3.6	0.9

Variable:	N	Attitudes Towards Evaluation: CDATE Mean	CDATE SD	ADATE Mean	ADATE SD	Total ATE Mean	Total ATE SD
High Apt	78	5.2	0.7	5.1	0.5	5.1	0.6
Low Apt	78	5.2	0.5	5.2	0.4	5.2	0.9
Total	156	5.1	0.6	5.1	0.5	5.1	0.6

Variable:	N	Attitudes Towards Mathematics: CDATM Mean	CDATM SD	ADATM Mean	ADATM SD	Total ATM Mean	Total ATM SD
High Apt	78	5.2	0.9	4.3	0.8	4.6	0.9
Low Apt	78	4.0	0.8	3.5	0.9	3.7	0.9
Total	156	4.6	0.9	3.9	0.9	4.3	1.0

*All means on a 1 to 7 scale

Table 10: INTERCORRELATIONS BETWEEN APTITUDE AND ATTITUDES TOWARDS TESTS, EVALUATION AND MATHEMATICS (N=156)

	APT	CDATT	ADATT	TATT	CDATM	ADATM	TATM
APT	1.0	.11	.21**	.18*	.43**	.53**	.47**
CDATT		1.0	.55**	.91**	.21**	.12	.21**
ADATT			1.0	.88**	.14	.35**	.29**
TATT				1.0	.22**	.27**	.30**
CDATM					1.0	.72**	.93**
TATM						1.0	.95**
CDATE	.02	.09	.01	.02	.01	-.05	.10
ADATE	.02	.07	.09	.01	-.03	.11	.07
TATE	.03	.09	.08	.02	.02	.06	.08

*=P<.05 **=P<.01

towards tests correlated with their attitudes towards math (r=+.30), but were uncorrelated with their attitude towards evaluation. Aptitude, however, was correlated to preparing teachers' attitudes towards tests at r=+.21 and their attitude towards mathematics at r=+.47, which is a relatively strong correlation in this context.

The correlations given in Table 10 are the same as those found in the studies reported above, as were the underlying factor structures of this group, which strongly cross-validate these results. The means and standard deviations given in Table 9 does not differ from the means and standard deviations we found for these scales in the studies reported above. Table 9 describes the means and standard deviations we have found for these attitudes in over 700 preparing teachers. Tests, therefore, are a very potent affective stimulus for a very large number of preparing teachers.

A wide variety of data that we have collected in several different contexts from over 300 preparing teachers clearly indicates that the two different "response arousal patterns" to the stimulus phase "To me tests are:" identified in study 1 above are highly correlated to "aptitude levels." For "low aptitude" preparing teachers, affective (highly emotional) adjectives tended to come up first, at a rate of roughly 3 to 1, as compared to cognitive adjectives, during the first two minutes of free association, whereas for "high aptitude" preparing teachers' cognitive adjectives came up first, in at a rate of roughly 4 to 1, as compared to affective adjectives, during the first 2 minutes of free association. This finding will be commented upon more fully

below.

Discussion and Conclusions

The findings in the six studies reported above need to be replicated, expanded and elaborated upon in a variety of ways. How pervasive and generalized our findings are is a question that needs to be answered. For example, we found that preparing teachers who had more than a cursory exposure to the sciences and/or psychology tended to have significantly more positive attitudes towards tests and mathematics than other preparing teachers. Unfortunately, we were not able to obtain SAT data on the preparing teachers in these studies, which is essential to clarifying effects and relations for a number of reasons related to both statistics and the literature on learning. Studies similar to the ones reported above need to be done on different groups such as middle and high school students and college students in a variety of different areas, as well as professionals in a variety of different areas. A variety of different variables should also be examined (such as field dependence and independence), and non-traditional assessment approaches such as free association and other such methodologies should be used.

One of the reasons we chose the semantic differential approach to measuring attitudes rather than other approaches is that the technique and the resulting data could be directly related to cognitive learning theory and cognitive theories of attention, perception, schema arousal, processing, memory (schema) structure formation, and responding. Other reasons were that the semantic

differential technique is less reactive and more versatile than other techniques and the items do not become dated in relatively short periods of time.

We believe that our 17 adjective-pairs for tests and evaluation performed much better and were of higher psychometric quality than the 15 adjective-pairs of the McCallon and Brown attitude towards mathematics scale. We, therefore, recommend that our 17 adjective-pairs be piloted for an attitude towards mathematics and an attitude towards statistics semantic differential, as such a scale does not currently exist. We also recommend that our 17 adjective-pairs be piloted for an attitude towards science semantic differential which is greatly needed, would complete this set of semantic differentials, and facilitate a great deal of much needed research in this area.

Given the results of the studies presented above, it would seem reasonable to ask what could be done about preparing teachers' attitude towards tests, given the importance of these attitudes to high quality education. Aside from using Wolpe's (1961, 1963) systematic desensitization technique in a wide variety of ways, one thing that can be done is to pay attention to the instructional sequence that is used to teach preparing teachers about tests and measurement. Nationally, preparing teachers tend to be taught educational tests and measurements by one of two very different instructional sequences which may be easily observed from a simple inspection of textbooks in this area.

The first of these two instructional sequences teaches preparing teachers testing and test theory first and then

descriptive statistics, correlation, and reliability and validity theory. This approach is sometimes called the classroom evaluation rather than educational tests and measurements approach. This approach to teaching educational tests and measurement can be traced to Tiegs (1926), who claimed that it was a practical, progessive, Deweyian approach that would motivate preparing teachers to learn educational tests and measurement by exposing preparing teachers to the most interesting and practical content in the discipline that they would use most frequently in their professional lives first rather than mathematical subject matter content. Tiegs claimed that preparing teachers had very negative attitudes towards mathematics and thus one could not teach them statistical concepts first, as such content would highly alienate preparing teachers immediately. Given Wolpe's theories, this is an extremely interesting claim in light of our data.

The Tiegs approach to teaching educational tests and measurement was blessed in the thirties by Tyler (1936) as the approach to teaching this subject matter content because it taught preparing teachers the management by (behavioral) objectives approach to teaching and education, which was much more important than educational tests and measurement and any disciplinary or cognitive learning theory concerns. By the sixties and seventies, the Tyler approach was espoused by behaviorists such as Robert Thorndike as being the only logical and common sense approach to teaching this subject matter content to preparing teachers. This approach to teaching educational tests and measurements to preparing teachers is the dominant approach currently.

The other approach to teaching preparing teachers educational tests and measurement is an older approach than the Tiegs, Tyler, behaviorist approach. This approach teaches preparing teachers statistical concepts first (variation and correlation), then reliability and validity, and then lastly testing and test construction theory. This approach was devised by E.L. Thorndike (1917) for the first formal course in educational tests and measurements taught in America which he taught at Columbia in 1914. Thorndike's rationale for his approach was that preparing teachers needed to learn an underlying conceptual framework first so that they could understand testing and test theory and its rational. The logical underpinnings for the Thorndikian approach was formalized by Smith (1936), also at Columbia, and anticipated Bruner and the structure of the discipline approach by 35 years. Cronbach (1970) and others have consistently argued for this second approach to teaching educational tests and measurements to preparing teachers. This approach is really a cognitive information processing approach to teaching educational tests and measurements to preparing teachers, whose theoretical underpinnings and operations are outlined in the work of Tulving (1972), Royer (1978), and Milligan (1983). This approach to teaching preparing teachers educational tests and measurements develops and cumulatively builds appropriate conceptual (subject-matter) schemas and a hierarchical, explanatory-interpretative framework in the learner for perceiving, decoding, elaborating, comprehending, encoding and encorporating the subject-matter content to-be-learned directly into long term semantic rather than episodic memory at

each point in the instructional process.

We compared the effectiveness of these two approaches to teaching preparing teachers educational tests and measurment in a fully cross Approach by Instructors by Aptitude level design (2x4x2). We found that preparing teachers in the Cognitive Approach achieved better and constructed better achievement tests than prpearing teachers in the Management by (behavioral) Objectives Approach at the .001 level of significance. We found that preparing teachers in the Cognitive Approach showed large positive changes in their attitudes towards tests, whereas preparing teachers in the Management by (behavioral) Objectives Approach showed no changes at all in their attitudes towards tests which became slightly more negative. We also found that low aptitude preparing teachers in the Cognitive approach performed as well as high aptitude preparing teachers in the Management by (behavioral) Objectives Approach. We found no instructor differences in the results obtained and that low aptitude preparing teachers in the Management by (behavioral) Objectives Approach performed the worst of all groups studied. These findings will be presented in fuller detail in another paper, but what needs to be explained in all of this data is the strong aptitude effects that we keep finding.

The differences between high and low aptitude preparing teachers attitudes towards tests may be explained in three different ways. These three different models are illustrative of the explanations that may be posited for the aptitude differences found on any of the variables in our studies. We are focusing on

the attitude towards tests variable as it is the most germane.

The first model is Kirland's (1971) view that the results observed are effects (and not causes) of preparing teachers' long history of experience with poor teacher made tests and evaluations in their K-12 years, which far outnumber their experiences ith professionally developed and validated standardized tests by a factor of 100 to 1 at a minimum. This very simple fact is one of the major reasons why Kirland, who sees testing as harmful, stresses a strong and highly concerted focus on the actual tests teachers construct and implement in their classroom, and teachers' knowledges of and ability to apply the principles of educational tests and measurement in their classroom, as a key variable to the improvement of the quality of education in this country.

The second explanation or model of the aptitude effects found in our data is the behaviorist view or model. This view says that high aptitude preparing teachers have more positive attitudes towards tests than low aptitude preparing teachers because high aptitude preparing teachers have had more rewards and successes as a consequence of the testing experience than low aptitude preparing teachers. Consequently, we must either diminish the rewards of testing, or improve the positive aspects of the testing experience for low aptitude students, or both. Given that this explanation or model is not only circular and full of hidden assumptions and contradicted by volumes of data from ETS and elsewhere, it begs the question.

Tests are currently a potent affective stimulus for all preparing teachers (if our findings are widely replicated) and

perhaps many, many others. Tests and testing, however, are one of the major ways and processes by which all human beings come to grips with actual and competitive realities, ineluctibly and necessarily. That tests are a potent affective stimulus for preparing teachers or anyone else is not a trivial problem, regardless of the numerous qualifications and caveats one might wish to add to these findings. Nor is the fact that preparing teachers have negative attitudes towards tests and testing a trivial problem. In a word, if tests are a highly potent affective stimulus for the citizens of this nation and those who teach them, then we are in extremely serious trouble competitively, and in terms of the world economy and global competition. Neither the behaviorist model nor its recommendations are real explanations or hold real answers to the questions our research and findings raise.

The third explanation or model of the aptitude effects in our data in a model tentatively and broadly outlined by Rothkoff and others in the seventies (Rothkopf, 1970). This model was first investigated by House (1966) and is a very cognitively oriented model. This model says that (a) aptitude is a surrogate variable, and that (b) "high aptitude" students behave and behave covertly in ways that are very different from "low aptitude" students. "High aptitude" students are more active learners and processors of information and experiences who actively elaborate and seek out relationships, inferences, and consequences, and continually reformulate and adapt their schemas based upon their highly active processing of experience. High aptitude students are better at finding masked relationships, overcoming the effects of set and

external enviromental factors and influences. They are better at spontaneously organizing and reorganizing their experiences and schemas, and finding and process structure (see Royer, 1978), and they do not passively accept or assimilate views, propositions, beliefs, attitudes, or systems "fed" to them by others via passive and receptive models of learning and teaching.

High aptitude students are self-directed, active, field independent, good at handling "knuckleballs," and highly constrained problem solvers with high meta-cognitive, meta-affective, and mathemagenic skills; namely, the very kind of teachers that we all say that we would like to have in our classroom (if we believe our rhetoric). In a word, the very kind of person that the current educational process in this country systematically discourages and seeks to modify behaviorally from daycare right on through to employment in our classrooms. We should be highly thankful and consider ourselves blessed that such behaviors are developmental characteristics that are extremely difficult to change or modify, as can be observed by going into many of the classrooms in our country and watching a large number of teachers who should be called courageous rather than the other adjectives that tend to be used in describning these teachers.

This cognitive view and explanation of the aptitude effect, we believe, is the better of the three views presented, and the one that is most consonant with and best supported by the experimental literature. The good news is that this very same empirical literature strongly supports the fact that with the right kind of instruction, educational experiences and academic enviroment, "low"

aptitude students improve significantly and begin to perform and behave like high aptitude students as our data shows. The bad news is that tests are currently a highly potent affective stimulus for preparing teachers (and most likely many, many others), and this is not a condition that can just be accepted as a fact and passively ignored, or we are going to be in extremely serious trouble competitively, and in terms of the world economy and global competition. There is a developmental, adjustment, and very hard reality factor in tests and testing that simply must be addressed and faced across the board concerning teachers' and everyone else's "attitudes towards tests" and "comfortableness with tests and the testing situation" that simply cannot be passively ignored, if we really wish to be creatively and dynamically competitive in a healthy, constructive, and balanced way.

References

Anderson, N. Integration theory and attitude change. Psychological Review, 1981, May, 78, 3, 171-201.

Blousfield, W.A. The occurence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 1953, 49, 229-240.

Campbell, D.T. and Fiske, D.W. Convergent and discriminant validation by the multi-trait-multimethod matrix. Problems in Human Assessment. Jackson and Mesnick (Eds.), McGraw-Hill, 1967, 124-132.

Cook W. and Selltiz, C. A multiple indicator approach to attitude measurement. Psychological Bulletin, 1964, 62, 36-55.

Cronbach, L. Essentials of Psychological Testing. Harper Row, New York, 1970.

Evans, J.E. Some characteristics of preparing and in-service teachers. Paper presented at the New England Educational Research Association Conference, Plainville, New York, Oct 26-28, 1983.

Dagenais, F. and Marascuilo, L. The effects of factor scores, Guttman scores, and simple sum scores on the size of F-ratios in an analysis of variance design. Multivariate Behavioral Research, 1973, 491-502.

Festinger, L. A Theory of Cognitive Dissonance. Harper-Row, New York, 1957.

Flanagan, J.C. The critical incident technique. Psychological Bulletin. Volume 51, No 4, July, 1954, 327-348. Prentice-Hall, Engelwoods.

Garskoff, B.E. and Houston, J.B. Measurement of verbal relatedness. Psychological Review, 1963, 70, 277-288.

Goslin, D.A. Teachers and Testing. Russell Sage Foundation, New York, 1967.

Glass, G.V., Peckham, P.D., and Sanders, J.R. Consequences of failure to meet assumptions underlying the analysis of variance and covariance. Review of Educational Research, 1972, Summer, 42, 3, 237-288.

Guilford, J.P. Psychometric Methods. McGraw-Hill, New York, 1965.

Harmond, H.H. Modern Factor Analysis. University of Chicago Press. Chicago, 1965.

Harris, C. Problems in Measuring Change. University of Wisconsin Press, Madison, 1967.

House, B. Discrimination of sysmetrical and asymmetrical dot patterns. Journal of Experimental Child Psychology, 1966, 3, 377-389

Johnson, P.E. Word relatedness and problem solving in physics. Journal of Educational Psychology, 1971, 55, 2, 84-88.

Kirkland, M.C. The effects of tests on students and schools. Review of Educational Research, 1971, 41, 4, 303-350.

Kniefel, D. and Dinham, S. A survey of educational measurment text-books. Journal of Educational Measurement, 1970, 1, 2, 63-71.

Likert, R. A technique for the measurement of attitude. In Public Opinion and the Individual, G. Murphy and R. Likert (Eds.), Harper, New York, 1937.

Marsullio, J. Approaches to teaching educational measurement to preparing teachers. Paper presented at the American Educational Research Association Conference, 1971, Chicago, 1-28.

Martin E. The encoding variablity hypothesis. Psychological Review, 1968, 75, 5, 421-441.

Mayo, R. Pre-Service Preparation of Teachers in Educational Measurement. U.S. Department of Health, Education, and Welfare, 1967.

McCallon, D. and Brown, J. A semantic differential for measuring attitudes towards mathematics. Journal of Experimental Education, 1971, Summer, 39, 4, 69-72.

McGuire, W.J. Nature of attitudes and attitude changes; in G. Lindsay and E. Aronson (Eds.) Handbook of Social Psychology. Addison-Wesley, Reading Ma., 1969.

Milligan, J. Schema leaning theory. Review of Educational Research, 1983, Spring, 53, 1, 197-207.

Neisser U. Cognitive Psychology. Cliffs, N.J., 1967.

Noll, V.H. Requirements in educational measurement for prospective teachers. School and Society, 1965, Septemeber, 17, 82, 88-90.

Norman, D. Memory and Attention: An Introduction to Human Information Processing, Wiley, New York, 1981.

Nunnally, J.C. Psychometric Theory. McGraw-Hill, New York, 1967.

Osborne, R. The preferential training needs record: a study of in-service training needs of teachers in the Atlanta Educational Service Area. Journal of Experimental Education, 1933, June, 19, 20, 37-49.

Osgood, C.E., Suci, G.J., and Tannenbaum, P.H. The Measurement of Meaning. University of Illinois Press, Urbana, 1964.

Peak, H. Attitude and motivation. Nebraska Symposium on Motivation. University of Nebraska Press, Lincoln, 1955.

Richards, J. A simulation study of the use of change measures to compare educational programs. American Educational Research Journal, 1972, 12, 3, 299-311.

Rothkopf E.Z. The concept of mathemagenic activities. Review of Educational Research, 1970, 40, 325-336.

Royer, F.I. Intelligence and the process of stimulus structure. Intelligence, 1978, 2, 11-40.

Shavelson, R.J. Some aspects of the correspondence between content structure and cognitive structure in physics instruction. Journal of Educational Psychology, 1971, 63, 2, 225-234.

Snider, J. and Osgood C. Semantic Differential Technique: A Sourcebook. Aldine, Chicago, 1969.

Stephens, J. Applied Multivariate Statistics for the Social Sciences. Lawerence Eribaum, Hillsdale N.J, 1986.

Smith O. Logical Aspects of Educational Measurement. Columbia University Press, New York, 1938.

Summers, G.F. Attitude Measurement. Rand-McNally, Chicago, 1970.

Tatsuoka, M.M. Multivariate Analysis. MacMillian, New York, 1988.

Tiegs, E. Tests and Measurements for Teachers. University of Chicago Press, 1929.

Thorndike, E.L. An Introduction to Educational Tests and Measurement. Columbia University Press, New York, 1917.

Tulving, E. Subjective organization in free recall of "unrelated" words. Psychological Review, 1962, 69, 344-354.

Tulving, E. Organization of Memory. Academic Press, New York, 1972.

Tyler. R. The function of measurement in improving instruction. Journal of Educational Research, 1936, 47-67.

Wople, J. The systematic desensitization of treatment neurosis. Journal of Mental and Nervous Disease, 1961, 132, 189-203.

Wople, J. Behavioral therapy in complex neurotic states. British Journal of Psychiatry, 110, 1963, 28-34.