Educational Tests and Measurements

Part I

The Assessment of Noncognitive Educational Outcomes

Reginald L. Jones

Cognitive objectives emphasize remembering or reproducing something which has presumably been learned, as well as objectives which involve the solving of some intellective task for which the individual has to determine the essential problem and then reorder the given material to combine it with ideas, methods, or procedures previously learned. Cognitive objectives vary from simple recall of material learned to highly original and creative ways of combining and synthesizing new ideas and materials. Virtually all educational objectives fall into this domain, and their use represents the basis for assigning grades, and determining the effectiveness of instruction.

Quite often instructors hold objectives in the *affective domain*, although student achievement of affective objectives, particularly at the college level, is infrequently used as a basis for grade assignment. Affective objectives are those which emphasize a feeling tone, an emotion, or a degree of acceptance or rejection. Affective objectives vary from simple attention to selected phenomena to complex but internally consistent qualities of character and conscience. Typically, affective objectives are expressed as interests, attitudes, appreciations, values, and emotional sets or biases.

A taxonomy or classification scheme has been developed for assessing educational objectives in the affective domain. However, the affective taxonomy seems less useful for framing instructional objectives than does the cognitive taxonomy.

It is assumed that an instructor may be interested in assessing student attitudes toward the course, field, or subject matter, or student values regarding some aspect of the field. Such assessments will not form the basis for grade assignment. Nevertheless, they can enrich teaching by familiarizing the instructor with a range of student values and perceptions which may be related to his/her course objectives. Some instructors may find particularly useful pre-course and postcourse studies of student responses to certain affective outcomes which they expect students to achieve during a semester, year, or indeed, during a three or four year course of study. Not infrequently, instructor familiarity with student responses to such assessments may form the basis for course or curriculum modifications to include a wider range of experiences, to emphasize different objectives, etc.

The remainder of this article will be devoted to brief descriptions of certain techniques which can be utilized to assess objectives in the affective domain.

PROCEDURES FOR ASSESSING AFFECTIVE OUTCOMES

There are, of course, a number of procedures which an instructor can use to assess attitudes towards his course, or to some subject matter related to it. Possible measures include open-ended questionnaires in which the instructor merely phrases a series of free response questions reflecting topics in which he has interest, or interviews, also open-ended, in which small groups of students are questioned about the attitudinal domain of interest. Both procedures have in common the fact that they are relatively easy to construct, and are open-ended.

There are a number of scales which have proved useful for attitudinal assessment. A scale is a set of verbal items to which an individual responds by expressing degrees to agreement or disagreement, like-dislike, etc. Three major types of attitude scales are: (1) summated rating scales (Likert), (2) equal appearing interval scales (Thurstone), and cumulative scales (Guttman).

Equal appearing interval scales and cumulative scales take some time and specialized skill to develop, and within the educational context, for the information yielded, probably do not justify the effort. Likert scales (and the Semantic Differential Scale to be described shortly) are relatively easy to construct, and moreover yield information which is as valid and reliable as that of the more sophisticated scales.

GENERAL CONSIDERATIONS

Regardless of the type of attitudinal measure chosen for use, several preliminary steps need to be considered. The first is a conceptualization of the domain. In practical terms this means developing a blueprint or outline of the components of attitudes to be assessed. If we are interested in assessing course attitudes for example, we might wish to make certain that all aspects of the course are covered. Thus, the blueprint might include the following components: (1) attitudes toward the lectures, (2) attitudes towards the films, (3) attitudes toward the field trips, (4) attitudes toward the textbook, (5) attitudes toward the assigned readings, and (6) attitudes towards the examinations. It is unlikely that any single course will include all of these components. However, if such were to be the case then adequate assessment of student attitudes means that each of these components would have to be taken into account.

One important by-product of the blueprint is its potential for yielding diagnostic information about attitudes toward a course. It is highly unlikely that students will hold uniformly high or low attitudes regarding all course components. The blueprint permits the instructor to know which aspects of the course are rated favorably or unfavorably, and consequently, those course aspects that may need to be modified.

A second consideration concerns the source of individual items. There are no specific rules to be followed here. The test developer's own armchair analysis,

previous tests, or anonymously completed essays from which attitudinal items are extracted are all possible sources of items. Once items are secured however, it is useful to have them checked for clarity by a colleague or some other person.

A third consideration, actually independent of questionnaire format, concerns respondent anonyomity. If valid responses are to be obtained, the student should be requested not to sign his name.

In the sections following, two procedures for assessing affective outcomes will be described — the Semantic Differential Technique and the Likert Scale. These procedures are easy to administer and score and, for most purposes, can be initiated and carried out by the instructor without the assistance of specialized personnel.

THE SEMANTIC DIFFERENTIAL

The Semantic Differential technique is used to measure generalized attitudes towards objects, persons, or concepts. Almost any concept of interest to an instructor can be used. For example, an English teacher might want to explore generalized attitudes toward poetry, toward essay writing, or toward literature. Instructors of geography, physics, and economics may wish to explore concepts related to these fields. Assessment areas are limitless.

The format of the Semantic Differential involves a rating scale of bipolar adjective pairs related to a concept of interest to the evaluator. Typical instructions, and an example of scale use in the evaluation of instruction, are given below:

TYPICAL INSTRUCTIONS

The purpose of this study is to measure the *meanings* of certain things to various people by having them judge them against a series of descriptive scales. In completing this form, please make your judgments on the basis of what these things mean to *you*. On each page of this booklet you will find a different concept to be judged and beneath it a set of scales. You are to rate the concept on each of these scales in order.

Here is how you are to use these scales:

If you feel that the concept at the top of the page is *very closely related* to one end of the scale, you should place your check-mark as follows:

fair—×—: ———: ———: ———: ———: —×—: unfair
fair———: ———: ———: ———: ———: ———: unfair

If you feel that the concept is *quite closely related* to one or the other end of the scale (but not extremely), you should place your check-mark as follows:

strong——: —×—: ———: ———: ———: ———: weak
strong——: ———: ———: ———: —×—: ———: weak

If the concept seems *only slightly related* to one side as opposed to the other side (but is not really neutral), then you should check as follows:

active——: ——: —×—: ——: ——: ——: passive
active——: ——: ——: —×—: ——: ——: passive

The direction toward which you check, of course, depends upon which of the two ends of the scale seem most characteristic of the thing you're judging. If you consider the concept to be *netural* on the scale, both sides of the scale *equally associated* with the concept, or if the scale is *completely irrelevant* or unrelated to the concept then you should place your check-mark in the middle space:

safe——: ——: ——: ——: ——: ——: dangerous

IMPORTANT: (1), Place your check-marks *in the middle of spaces*, not on the boundaries:

——: —×—: ——: ——: ——: ——:

(2) Be sure you check every scale for every concept—*do not omit any*,

(3) Never put more than one check-mark on a single scale.

Sometimes you may feel as though you've had the same item before. This will not be the case, *so do not look back and forth* through the items. Do not try to remember how you checked similar items earlier in the scale. *Make each item a separate and independent judgment*. Work at fairly high speed. Do not worry or puzzle over individual items. It is your first impressions, the immediate "feelings" about the items, that we want. On the other hand, please do not be careless, because we want your true impressions.

THE LABORATORY

1. important——: ——: ——: ——: ——: ——: unimportant
2. clarified the subject ——: ——: ——: ——: ——: ——: confused the subject
3. meaningless ——: ——: ——: ——: ——: ——: meaningful
4. interesting ——: ——: ——: ——: ——: ——: boring
5. should be given more time ——: ——: ——: ——: ——: ——: should be given less time
6. equipment inadequate ——: ——: ——: ——: ——: ——: equipment adequate
7. supplies adequate ——: ——: ——: ——: ——: ——: supplies inadequate

Above is given an example of use of the Semantic Differential in evaluation of the Laboratory.

It is apparent that, using the above format and similar or other descriptive adjectives, it will be possible to assess attitudes toward a number of matters, e.g. The Field Trips, The Examinations, The Lectures, The Value of This Course, etc.

Scoring and evaluation. Once students have completed the Semantic Differential, a number of analyses are possible. The simplest is to provide a tally of responses to each of the seven scale points and to compute the mean (simple average) or median for each item and for the full scale. Consider this example. Thirty students enrolled in a Chemistry Laboratory responded as follows:

	7(High)	6	5	4 (Neutral)	3	2	1 (Low)	Mean
1. important-unimportant	11	8	6	3	2	-	-	5.77
2. clarified the subject-did not clarify the subject	2	4	4	10	4	4	2	4.00
3. meaningful-meaningless	6	7	5	9	1	1	1	5.03
4. interesting-dull	5	8	10	7	-	-	-	5.37
5. should be given more time-should be given less time	-	-	-	-	10	10	10	2.00
6. equipment adequate-equipment inadequate	-	15	11	4	-	-	-	6.37
7. supplies adequate-supplies inadequate	-	-	-	-	6	7	8	9	2.33

It will be noted that all responses are rearranged from positive to negative in the above table even though they were randomly placed in the original form —i.e some positive stems placed at the left pole, for example "important," "interesting", "supplies adequate", etc., and some negative stems placed at the left pole, e.g., "meaningless," "equipment inadequate." Random placement of stems increased the likelihood that the student will respond to each item independently, minimizing generalized responses independent of item content.

Inspection of frequencies provided in the table can provide useful information about evaluation of the laboratory. It can be seen, for example, that most students evaluate the laboratory postively on the whole (item 1) and that it is perceived to be "meaningful" (item 3) and interesting" (item 4). Howevei, virtually all students agreed that more time should be given to the laboratory (item 5) and that the supplies were inadequate (item 7).

Other more sophisticated techniques for analysis of responses are possible, but the tallying of responses has the advantage of ease of application and as well yields information which can be used immediately in course (and oi concept) diagnosis and modification.

LIKERT SCALE

The Likert Scale, widely used, and probably familiar to most instructors, involves subject responses to a four or five point scale (usually ranging between strongly agree and strongly disagree) with respect to a given attitudinal item. With respect to developing item content, the general considerations suggested for construction of attitude scales (particularly development of the test bluerprint) will apply to the construction of Likert Scales. However, an additional consideration concerns the phrasing of individual attitude test items, certain rules for which are given below:

1. Avoid statements that may be interpreted in more than one way.
2. Avoid statements that are irrelevant to the object under consideration.
3. Select statements that are believed to cover the entire range of the scale of interest.
4. Keep the language of the statements simple, clear, and direct.
5. Statements should be short, rarely exceeding 20 words.
6. Each statement should contain only one complete thought.
7. Statements containing universals such as all, always, none, and never often introduce ambiguity and should be avoided.
8. Words such as only, just, merely, and others of a similar nature should be used with care and moderation in writing statements.
9. Whenever possible, statements should be in the form of simple sentences rather than in the form of compound or complex sentences.
10. Avoid the use of words that may not be understood by those who are to be given the completed scale.
11. Avoid the use of double negatives.

Examples of Likert items are given below. It is apparent that adjectives used in assessing student response to The Chemistry Laboratory, given as an example of use of the Semantic Differential in the evaluation of instruction, can be used also in a Likert format. As will be recalled, there were adjectives such as important-unimportant, clarified the subject-did not clarify the subject,supplies adequate-supplies inadequate, should be given more time-should be given less time, etc.

These adjectives could be used to phrase the following Likert itmes:

1. The Laboratory was an important part of the course.
   A. Strongly Agree B. Agree C. Uncertain D. Disagree E. Strongly Disagree

2. The Laboratory helped to clarify the subject matter.
   A. Strongly agree B. Agree C. Uncertain D. Disagree E. Strongly Disagree

3. The Laboratory supplies were adequate.
   A. Strongly agree B. Agree C. Uncertain D. Disagree E. Strongly Disagree

4. There should be more time alloted for laboratory work.
   A. Strongly Agree B. Agree C. Uncertain D. Disagree E. Strongly Disagree

For most course evaluation purposes, as with the Semantic Differential, frequencies of response to each alternative for each question need only be summarized and this information used in course (or concept) diagnosis and modification.

SUMMARY

A discussion of the assessment of certain affective outcomes of instruction (e.g. course attitudes, course concepts) and a description of two procedures for the appraisal of affective outcomes were presented. The procedures have the advantage of ease of construction and interpretation and thus can be easily adapted by an instructor for his/her particular needs.