DOCUMENT RESUME

ED 135 808 95 TM 005 901

AUTHOR Brzezinski, Evelyn; Ahn, Unhai
TITLE Program To Operationalize a New Training Pattern for Training Evaluation Personnel in Education. Final Report. Part A - Report on Development of Self Assessment of Evaluation Skills.
INSTITUTION Ohio State Univ., Columbus. Research Foundation.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
BUREAU NO 09039
PUB DATE 30 Jun 73
GRANT OEG-0-71-1051
NOTE 136p.; For related documents, see TM 005 902-904 ; Best Copy Available

EDRS PRICE MF-$0.83 HC-$7.35 Plus Postage.
DESCRIPTORS Educational Needs; *Educational Programs; *Educational Researchers; *Evaluation; Graduate Study; Higher Education; Measurement Techniques; National Surveys; *Self Evaluation; *Skills; Test Validity
IDENTIFIERS *Ohio State University Model Training Project; *Self Assessment of Evaluation Skills Instrument; Universe of Evaluation Competencies

ABSTRACT

The Model Training Project (MTP) based at The Ohio State University had as its major purpose the development of a training program which would, when implemented, meet the total range of evaluation training needs within the member agencies of the MTP consortium. One of the first jobs for members of the MTP staff was to try to define all the competencies required collectively by all the members of an evaluation team. That definition of skills and knowledge was called the Universe of Evaluation Competencies (UEC). It was recognized that the UEC must be validated by empirical studies, so the Self Assessment of Evaluation Skills (SAES) instrument included herein, was developed. The UEC was divided into ten major sections, each of which represented, in the view of its authors, a necessary component in the building of an evaluation profession. Skill areas were defined for the ten major sections; further, competencies were specified for each of the skill areas. In July, 1972, representatives from the MTP consortium agencies critigued the UEC and SAES. Later, consultants worked with the SAES staff in further revision of the SAES. The 234-item instrument which resulted was to be used by practicing evaluators in answering three questions: (1) How competent are you in this skill? (2) How important is this skill to the successful completion of your duties? and (3) How interested are you in performing this skill? The SAES pilot test (including methods and procedures used, results obtained, and conclusions drawn from the results) is also described. (Author/MV)

PROGRAM TO OPERATIONALIZE A NEW TRAINING PATTERN FOR TRAINING EVALUATION PERSONNEL IN EDUCATION

Final Report
Project Number 09039
Grant No. OEG-0-71-1051
National Institute of Education

Part A - Report on Development of
Self Assessment of Evaluation Skills

Prepared by:
Evelyn Brzezinski
Unhai Ahn

June 30, 1973

The Ohio State University Research Foundation
1314 Kinnear Road
Columbus, Ohio 43212

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

INTRODUCTION

One of the "close down products" of the Ohio State University-based Model Training Project requested by NIE personnel was a wrap-up of the current phase of development on the Self Assessment of Evaluation Skills (SAES) instrument, with documentation and reporting of all work completed. That documentation and reporting are the purpose of the report which follows.

Numerous persons have been involved with SAES since its inception, and their contributions to the work reported below are acknowledged at this time. Dr. Mary Anne Bunda and Dr. Daniel Stufflebeam should be recognized for the initial conceptualization of the SAES study, for the development of the SAES instrument, and for continued interest in the SAES project. Students from the Model Training Project who have been actively involved in the SAES study are Unhai Ahn, Phyllis Falk, and Andrea Lash. Specifically, the correlational analysis of the SAES data reported below was completed by Ahn; the follow-up of nonrespondents to the SAES instrument was conducted by Falk; and the item contamination study was completed by Lash. Nancy Riddle, Sharon Kelley, and Jennifer Hixson have been able secretaries to the SAES staff, and their help throughout this study is appreciated.

BACKGROUND OF THE STUDY

The Model Training Project (MTP) based at The Ohio State University had as its major purpose the development of a training program which would, when implemented, meet the total range of evaluation training needs within the member agencies of the MTP consortium. (Readers are referred to the document by Hilderbrand, et al., in this final report series which describes that consortium.) If the MTP's goal was achieved, staff members felt that this newly-developed program would provide one model of the types of evaluation training needed throughout the field of education.

During early Winter, 1971, the MTP staff held a retreat to concentrate on project development. A concern felt by most staff members was that training in evaluation should not become a set curriculum for all students. Rather, the staff members felt that some process which would allow individual program planning would better fit the structure of the MTP.

One of the major premises of the MTP was that evaluation is a team activity, and therefore persons must be trained to fill different positions on such a team. It follows, then, that persons must be trained differently, depending on their positions on a team -- i.e., depending on their evaluation role."

A problem exists, however, in that evaluation roles are presently at a primitive stage of definition. The skills and knowledge needed by persons who perform different evaluation roles are not widely recognized; even among those persons writing in this area, there is no consensus of just what it takes to be a "competent evaluator." Therefore, one of the first jobs for members of the MTP staff was to try to define all the competencies

required collectively by all the members of an evaluation team. That definition of skills and knowledge is called the Universe of Evaluation Competencies (UEC). Mainly the result of a conceptual effort, it was recognized that the UEC must be validated by empirical studies. The Self Assessment of Evaluation Skills (SAES) instrument was developed for that validation purpose.

The UEC was divided into ten major sections, each of which represented, in the view of the authors of the UEC, a necessary component in the building of an evaluation profession. These ten sections were: 1) Administrative Leadership in Evaluation, 2) Research in Evaluation, 3) Development of Evaluation Methodology, 4) Instruction in Evaluation, 5) Implementation in Evaluation, 6) Development of Evaluation Systems, 7) Diffusion of Evaluation Developments, 8) Use of Evaluation in Educational Leadership, 9) Use of Evaluation in Teaching, and 10) Use of Evaluation in Curriculum Development. Skill areas were defined for the ten major sections; further, competencies were specified for each of the skill areas. It is from this third level of the UEC that the SAES instrument was developed.

The UEC was checked for completeness against pertinent past research (e.g., the AERA Task Force on Research Training data, the Oregon Studies data). Each time a competency was found in that research which was not included in the UEC, it was added to the UEC. In addition, one of the researchers in this field of investigation (Blaine Worthen, the director of the AERA Task Force) examined the UEC to determine its comprehensiveness. After these two content validation steps, MTP staff felt relatively comfortable about the inclusiveness of the set of competencies listed in the

UEC, and also felt that the UEC and the related SAES instrument could serve as a basis for empirical research and program development.

Three uses were to be made of the SAES instrument. First, and most important for the purpose of validating the UEC, a national survey of members of the evaluation profession was to be conducted. Second, students in an evaluation training program could check their level of competence in certain skill areas with SAES. Third, learning experiences (e.g., classes, simulations) could be described on the basis of which skills from SAES they were designed to transmit.

When the instrument was first written, it contained approximately 300 items. The items were grouped according to the ten major sections of the UEC. Respondents (MTP staff and some students) were asked to rate their level of competence in each skill on a scale from 1 (no competence) to 5 (expert in the field).

The first revision of the instrument included a format change and clarification of ambiguous items. The titles of the ten UEC sections were dropped to lessen the possibility of bias in responses. The instrument was divided into three content equivalent forms of 113 items each (with some overlap of items among forms). This was done for purposes of practicality; it was hoped that shortening the instrument would increase the response rate in the national survey of practicing evaluators mentioned earlier. Also at this time, separate directions and answer sheets were prepared, corresponding to the three planned uses of SAES, for 1) practicing evaluators, 2) students, and 3) learning experiences.

After this revision, two groups of persons responded to the instrument in relation to their level of competence for each of the skills listed in

all three forms of the instrument. The two groups were 1) all students in an evaluation class at Ohio State University, and other MTP students not in the class, and 2) leaders in the field of evaluation who comprise the Working Group on Evaluation. A Q-factor analysis was performed on the data to see if the instrument would differentiate between students and practicing evaluators. The results of that analysis showed, in fact, perfect discrimination. Specifically, the first two factors differentiated the two groups of individuals perfectly with eigen values of 5.85 and 2.45, respectively. The first factor accounted for 20% of the variance and the second factor accounted for an additional 9% in an orthogonal space.

Also at this time, the three forms of the SAES instrument were analyzed to determine the equivalency of scores among the three forms and the internal consistency within the three forms. The results of that analysis were as follows:

Form 1A mean response - 1.96 (on a scale from 1-5)
Split-half reliability - .99

Form 1B mean response - 1.96 (on a scale from 1-5)
Split-half reliability - 1.00

Form 1C mean response - 1.87 (on a scale from 1-5)
Split-half reliability - .99

In July, 1972, representatives from the MTP consortium agencies critiqued the UEC and SAES. Several concerns about SAES were raised, prompting the MTP staff to engage consultants to work with the SAES staff in further revision of the instrument. The SAES instrument which resulted from these revisions was a 234-item instrument which was to be used by practicing evaluators in answering three questions: 1) How competent are you in this skill? 2) How important is this skill to the successful completion of your duties? and 3) How interested are you in performing

this skill?

The two new response scales, importance and interest, were added because the MTP staff and consortium representatives were hesitant to say that just because evaluators presently practicing in the field have certain competencies in certain skills, a training program should be developed on the basis of that data alone. The two new scales were developed to allow more information to be available to the MTP staff and other curriculum developers.

This version of SAES was administered to a small group of students and incumbent evaluators and was found to be relatively free of ambiguities. The instrument was judged, then, to be ready ror pilot testing. In the remainder of this report, that pilot test (including methods and procedures used, results obtained, and conclusions drawn from the results) is described.

PROCEDURES

The original purpose of the pilot test was to test the usefulness (e.g., applicability and understandability) of the SAES instrument on a small group of practicing evaluators before the full national survey was undertaken. Also, the SAES staff hoped to learn two additional things: 1) would the data received from the pilot test be amenable to the analyses that were planned for the full survey, and 2) what percentage of returned questionnaires might reasonably be expected for the full survey. In order to answer those questions, several different methods and procedures were used. A description of the procedures is included in this section of the report.

General Procedures

In this section, general procedures relating to the distribution of the SAES instrument are described. Included are the population and sample for the pilot test, questionnaires and cover letter used in the pilot test, data collection, and statistical treatment of the data. Following this section, descriptions are given of additional substudies which were undertaken in conjunction with the pilot test in order to answer specifically certain questions noted above.

Population and Sample. The population of concern in this study was all incumbent educational evaluators. It was uncertain how much of this experimental population was accessible. Four sources were used to complete a directory of practicing evaluators; approximately 2,500 persons were identified through these sources. The SAES staff believed that

these names represented a large enough proportion of the experimental population that the listing of names could, for all practical purposes, be considered complete.

The four sources which were used to identify incumbent evaluators are listed below.

1) The membership roster from Division H (School Evaluation and Program Development) of the American Educational Research Association

2) The mailing list for the interfacer, a newsletter published by the Model Training Project

3) Two lists of evaluation consultants compiled by Robert Stake

4) Personal address files of several evaluation specialists

In the full survey, it was planned that all persons in the identified population would be sent a SAES instrument. In the pilot test, however, a 10% sample was deemed sufficient to answer the questions of interest. Accordingly, a stratified random sample of 252 persons was drawn, with address (either home or office) being the stratifying variable. Of the 252 persons in the sample, 132 had office addresses and 120 had home addresses; this was the proportion of office and home addresses found in the experimental population.

Questionnaires and Cover Letter. As previously discussed, the SAES instrument was already developed and ready for use (a copy may be found in Appendix A of this report). Other materials which had been prepared previously and needed only minor revisions to be sent with the instrument were the Directions for Completing the Self Assessment of Evaluation Skills Instrument, the Answer Sheet, the Item Critique Form, and the Role Description Questionnaire (a copy of each is included in Appendix B).

The Role Description Questionnaire contained items intended to gather personalogical data about questionnaire respondents. Questions centered on the respondent's educational background and his employment responsibilities. One purpose of the pilot test was to validate the 24 role descriptions listed in the Role Description Questionnaire. It was hoped that in the full survey a checklist-type questionnaire could be developed to gather this personalogical data.

The cover letter accompanying the SAES and other instruments was a multilithed letter with each person's name and address typed on the letter. The letterhead of the stationery read "Commission for the Study of the Evaluation Profession." This Commission was organized by the SAES staff for two purposes: 1) to provide on-going critiques of the instruments, procedures, etc., to be used in both the pilot test and the full survey, and 2) to provide credibility for the study through the individuals listed as members of the Commission. Readers are referred to the Bunda paper (which appears later in this report) for a more complete description of the work of the Commission.

Data Collection. The initial mailing of cover letter, SAES, and other materials was made on November 10, 1972. Three weeks later, on November 30, 1972, a follow-up reminder was sent. As will be explained in a later section of this report, half of the nonrespondents at this time received a postcard follow-up and the other half of the nonrespondents received a follow-up letter with all the instruments and enclosures that had been included in the initial mailing. A second follow-up reminder, this time a postcard for all nonrespondents, was sent on December 15, 1972. January 12, 1973, was the date chosen as the cut-off, beyond which returned

questionnaires were not included in the analyses.

Each packet of materials was numbered so that returned questionnaires could be checked off a master list. As will be discussed later, one of three types of return envelope was included in each packet, the three types of envelope being stamped, business reply, and unstamped. (Copies of all follow-up correspondence may be found in Appendix C.)

Three Substudies

In the next three sections, specific procedures are described relating to three substudies undertaken to answer these questions:

1) How could the response rate be increased without increasing the cost of the study?

2) Are the SAES items clear?

3) Why did some persons fail to respond to the request to complete the SAES instrument?

Procedures for Mailing Costs Substudy. As in any mailed questionnaire study, the SAES staff was concerned about maximizing response rates. In most cases, the validity of the results of a questionnaire study is only as good as the response rate, i.e., the higher the response rate, the more confidence one can place in the results. There are many techniques which researchers employ to try to attain high response rates, and some of the techniques (such as personal or telephone follow-up) do seem to be effective in increasing return rates.

Unfortunately, some of the more effective techniques are quite expensive. In working with a pilot test sample of 252, the SAES staff could have felt relatively free to do whatever they could to achieve high response rates, regardless of the cost. Since a primary purpose of the

pilot test was to estimate the response rate the full survey might achieve, however, the SAES staff could do nothing in the pilot test that they would not do in the full survey. With the number of potential respondents set at 2,500 in the full survey, cost did become a factor. In the pilot test, therefore, two variables were investigated to determine the most cost-effective means for achieving a high response rate.

The two variables included in this substudy were mode of follow-up and format of return envelope. A completely crossed 2 x 3 factorial design (two levels of follow-up, three levels of return envelope) was employed. Pilot test respondents were randomly assigned to the six cells, with the constraint that the population proportion of respondents with home or office addresses was maintained in these six cells. (In actuality, there were 36 cells in the design. A third classification variable relating to the item critique substudy was included in the experimental design; there were six levels of that variable.)

The two levels of follow-up were postcard and letter (with SAES and all other materials from the initial mailing included for a second time). This dichotomy was studied for the first follow-up contact only; at the time of the second follow-up, all nonrespondents received a postcard reminder.

The three return envelope formats were stamped, business reply, and unstamped. Except for the postage, all envelopes were the same. The effect of this variable was studied through the first follow-up as well.

Procedures for Item Critique and Item Contamination Substudies. Persons who had critiqued the SAES instrument questioned the appropriateness of items containing two verbs in a single item. These items forced

respondents to produce one rating for what might be two separate skills.

In order to determine if information on respondents' skills was lost in the double verb items (because respondents may have to average the ratings for each skill or respond to one skill only), a substudy was included in the SAES pilot test. The purpose of the substudy was to see if responses to double verb items (implying two skills) differed from responses to each of two single verb items measuring the same skills.

All double verb items were noted in the SAES instrument and reviewed by the SAES staff. Items whose verbs could imply two separate skills (e.g., to design and administer an evaluation), or two levels of one skill, were chosen for the study. Twenty-eight such items were chosen for inclusion.

Each double verb item was then rewritten into two separate items (so respondents could answer each item separately). These new, single verb items were reviewed by the SAES staff for clarity and appropriateness.

Samples for three cells were randomly drawn from the population, which at this time contained the approximately 2,500 names minus the 252 persons in the pilot test sample. A total of 120 persons was drawn and 40 persons were randomly assigned to each cell. The population's proportion of home and office addresses was maintained in the sample and in the cells.

Each respondent received the Role Description Questionnaire, a stamped return envelope, and a set of SAES items. The packets were mailed November 17, 1972.

Group 1 received all new single verb items (N = 56) and 8 single verb items randomly chosen from the original SAES for a total of 64 items.

Group 2 received 1/2 of the new single verb items (N = 28, one item randomly chosen from each pair of items) and the same 8 single verb items from the original SAES for a total of 36 items.

Group 3 received all 28 double verb items and the same 8 single verb items from the original SAES for a total of 36 items.

The eight single verb items chosen from the original SAES were included so responses to these items could be compared among the three groups. It was assumed that mean responses to these items would be (statistically) the same for all three groups.

Subjects who had not returned their answer sheets were sent follow-up reminders on December 7, 1972. These follow-ups included the same material that was in the original mailing. (The items included in this substudy, and all materials sent to the respondents in this sample, are included in Appendix D.)

In addition to this substudy, each person in the pilot test sample of 252 critiqued a portion of the SAES items or the Role Description Questionnaire. The total sample of 252 persons was divided into six groups: five groups critiqued 50 items each (for clarity and ease in responding), and one group critiqued the Role Description Questionnaire. For the 234 SAES items, respondents answered two Likert-type items for each SAES item. The descriptors for the clarity scale were:

1) The item was completely understandable.

2) The item was moderately clear.

3) The item was moderately ambiguous.

4) The item was completely ambiguous.

The descriptors for the ease of answering scale were:

1) The item was easy to answer.

2) The item was moderately easy.

3) The item was moderately difficult.

4) The item was difficult to answer.

The critique of the Role Description Questionnaire was conducted by answering five open-ended items about the questionnaire.

Procedures for Nonrespondent Bias Check. In order to assess possible differences between respondents and nonrespondents, a telephone follow-up of nonrespondents was conducted. Approximately fifteen percent of the nonrespondents (N = 22) was randomly selected to be included in the follow-up study. This fifteen percent of nonrespondents was compared with approximately fifteen percent of the respondents who returned usable questionnaires.

In the pilot study, the sample had been divided into several subgroups. Some of the sample had home addresses while some had office addresses. In each of these two groups, the sample was further divided by types of return envelope: some persons received stamped return envelopes, some received business reply envelopes, and some received envelopes with no return postage. A proportionate number of respondents and nonrespondents were sampled from each of the subgroups. The number of members in the subgroups was often not conducive to proportionate sampling within the restrictions of a fifteen percent sample. For cases in which an additional member was needed to complete the fifteen percent sample, the subgroup from which the additional member was drawn was randomly selected.

The respondents and nonrespondents were compared on the following six variables:

1) position title

2) role description

3) employing institution/department

4) length of time in present position

5) highest degree

6) date of highest degree

In addition to the six variables listed above, another variable was investigated. In order to obtain an indication of the reasons for their failure to respond to the questionnaire, the nonrespondents were queried in regard to this issue.

In the next section, general results as well as results for each of the substudies described above are presented. The statistical treatment of the data is discussed as an introductory segment to each set of results.

RESULTS

In this section of the report, results from the various substudies and other analyses are given. In addition, the statistical techniques used in each analysis are listed before the results are described.

General Results

Thirty-one percent (N = 77) of the 252 SAES instruments mailed to respondents were returned to the SAES staff with completely usable responses. An additional eight questionnaires were partially completed and returned, while 21 were either undeliverable or returned blank.

The Role Description Questionnaires returned by the 77 persons who responded completely to the SAES instrument provided information about the educational and professional experience of the respondents. Table 1 shows the distribution of respondents among the various categories contained in the questionnaire. Professional experience data relates to the respondents' present position. Because respondents could choose more than one descriptor in certain categories (e.g., university administrator and university professor), the figures below do not always sum to 77.

Table 1

Description of Respondents
to SAES Instrument (N = 77)

Present Job Title	
Administrator: School building - 1	Administrator: College or univ. - 14
Teacher: Elementary or secondary - 1	Professor: College or univ. - 26
RDE Spec[1]: School building - 1	RDE Spec: College or univ. - 5
Administrator: School district - 6	Administrator: R & D Lab - 2
RDE Spec: School district - 5	Specialist: R & D Lab - 4

(Table 1 continued)

[1] Research, Development, or Evaluation Specialist

Table 1 (continued)

Description of Respondents
to SAES Instrument (N = 77)

Present Job Title (continued)

Specialist[2]: School district - 1
Administrator: State dept. of educ. - 1
RDE Spec: State dept. of educ. - 4
Other administrator - 8
Other RDE specialist - 3

Present Role Description

Administrator: Research - 11
Administrator: Evaluation - 9
Administrator: Development - 1
Other administrator - 19
Professor: Research courses - 9
Professor: Evaluation courses - 3
Professor: Development courses - 6
Other college professor - 16
Teacher: Elementary or secondary - 1
Specialist: Evaluation design - 11
Specialist: Evaluation implementation - 7
Consultant: Evaluation - 10
Educational program auditor - 2
Other consultant - 2
Product developer - 2
Program developer - 2
Researcher - 12
Statistician/data analyst - 4
Psychologist - 1
Specialist: Diffusion/dissemination - 3
Evaluator of federal projects - 8
Counselor - 2
Director of testing program - 2
Institutional researcher - 2
Other - 1

Present employing Agency/Department

University/Education - 28
University/Psychology - 1
Univ./specially funded project or center - 8
School system/elem. school - 1
School system/sec. school - 1
School system/evaluation, accountability, and/or planning and research - 8
School system/special education - 1
School system/testing, student information - 1
School system/guidance, pupil personnel services - 2
Lab or Center/specially funded projects - 6
State dept. of educ./research, evaluation, planning - 3
State dept. of educ./special projects - 2
Other - 9

Length of Time in Present Position

Less than 1 year - 13
1 - 2 years - 22
3 - 5 years - 20
6 - 10 years - 8
More than 10 years - 4

Degree Required for Present Position

Bachelor's - 3
Master's - 21
Master's plus - 2
Doctorate - 39

(Table 1 continued)

[2] Specialist at district level, e.g., psychologist

Table 1 (continued)

Description of Respondents
to SAES Instrument (N = 77)

Certification Required for Present Position		
None - 44	Supervision - 2	Social Work - 1
Teaching - 5	Administration - 4	Other - 4

It would appear from looking at Table 1 that the sample was over-represented with persons from colleges and universities. This is understandable, since it is likely that the vast majority of AERA members (the primary source of names in the SAES staff's directory of evaluators) are college or university personnel. Before the full survey could have been undertaken, it would probably have been necessary to augment the directory with the names of (possibly numerous) evaluators from settings other than universities. (This point is discussed further in the "Limitations of this Study" section later in this report.)

Results of Correlation Analysis

The 234 SAES items which were sent to the pilot test sample were developed from the Universe of Evaluation Competencies, which has been described in an earlier section of this report. The 234 items were grouped on an *a priori* basis into eight categories of items; each category was thought to represent a distinct skill area necessary for the successful completion of evaluation activities. One of the purposes of the full survey was to test empirically this logical grouping of items through the use of factor analytic techniques. Because of the small number of responses to the pilot test (N = 77), it was impossible to factor analyze the pilot test results, but the SAES staff wanted to

have some idea (even though it might be tenuous) of the validity of the eight logical groups. For this reason, correlation analyses of the pilot test data were performed.

Two different correlation analyses were performed. One was undertaken to determine inter-item correlations among the 234 items. The BMD X84 program was used for this analysis. The second analysis was concerned with inter-scale correlations (i.e., the intercorrelations among the competence, importance, and interest scales). The BMD 03D program was used for this analysis.

Inter-item correlations. The eight logically derived areas or categories of skills comprising the SAES instrument are as follows: 1) knowledge of innovation in evaluation, 2) public relations, 3) data processing, 4) educational measurement, 5) evaluation administration, 6) relating evaluation to relevant disciplines, 7) communications, and 8) research design analysis. A list of the items corresponding to each skill area may be found in Appendix E.

For this analysis, items were grouped according to the category in which they were placed on *a priori* grounds. An inter-item correlation matrix was computed for each of the eight categories of items. Then the items were put back in numerical order (i.e., they were no longer grouped according to the eight categories) and another inter-item correlation matrix was formed. Using a table of critical values for the correlation coefficient (Glass & Stanley, Statistical Methods in Education and Psychology, p. 536), all correlations which were significant at the .01 level of confidence were noted. Table 2 shows the percentage of significant correlations in each category as well as the percentage of significant correlations in the total SAES instrument. That is, the percent

of significant correlations between an item and all other items not in its logical group is the number shown in the "Total" entry.

Table 2

Percent of Significant Inter-item Correlations

Category	Number of Items in Category	Percent of Significant Correlations
1. Knowledge of innovation in evaluation	46	72.8
2. Public relations	8	66.7
3. Data processing	11	87.2
4. Educational measurement	34	65.4
5. Evaluation administration	50	76.4
6. Relating evaluation to relevant disciplines	12	84.8
7. Communications	22	44.5
8. Research design analysis	51	68.2
Total	234	40.2

As shown in Table 2, items within each category appear to have higher intercorrelations than the items between categories. The cohesiveness of the items was most evident in categories 1, 3, 5, 6, and 8. These five categories had more significant correlations ($p < .01$) and categories 2 and 4 had more significant correlations ($p < .05$) than the intercorrelations among the total set of items (the 40.2% figure noted above). The percent of correlations in category 7 was not significantly different from the total group.

Inter-scale correlations. Because it was important to determine the independence of the three response scales (competence, interest, and importance), an inter-scale correlation matrix was computed for the total set of items and for the eight logically derived categories of items. The results of that analysis are shown in Table 3.

Table 3

Average Inter-scale Correlations

Category	Number of Items in Category	r_{ci}	r_{cm}	r_{im}
1. Knowledge of innovation in evaluation	46	.55	.46	.64
2. Public relations	8	.52	.33	.47
3. Data processing	11	.54	.35	.47
4. Educational measurement	34	.53	.43	.60
5. Evaluation administration	50	.56	.52	.60
6. Relating evaluation to relevant disciplines	12	.56	.49	.62
7. Communications	22	.52	.43*	.61*
8. Research design analysis	51	.55	.43	.59
Total	234	.54	.45	.59

NOTE: c = competence, i = interest, m = importance

* The difference between .43 and .61 is significant at the .05 level, but because of the number of correlations computed, this might be expected by chance.

In general, there were high correlations among the three scales. The correlation between the interest and importance scales (r_{im}) was the highest, with an average of .59. The correlation between the competence and interest scales (r_{ci}) was the second highest, .54. The correlation between the competence and importance scales (r_{cm}) was the lowest, .45. However, there was no significant difference among these three correlations. This overall trend was also evident when items were grouped for each category, with the exception of categories 2 and 3.

It is difficult to draw any clear conclusions from the above results. It seems safe to say, however, that the three scales are not really independent. Before the full survey would be conducted, it would be wise to determine if all three scales are actually needed.

In Appendix F, inter-scale correlations for all items may be found.

Results of Item and Role Description Questionnaire Critique

As previously discussed, the 252 persons in the pilot test sample were randomly assigned to six groups for the purpose of critiqueing the 234 SAES items and the Role Description Questionnaire. Responses within these six groups were as follows:

Group	Responses	Percent
1 (critique of items 1-50)	11	26.2
2 (critique of items 51-100)	14	33.3
3 (critique of items 101-150)	9	21.4*
4 (critique of items 151-200)	17	40.5*
5 (critique of items 201-234)	9	21.4*
6 (critique of Role Description Questionnaire)	11	26.2

* The difference between 21.4% and 40.5% is significant at the .05 level. That should not be a limitation, however, of the results reported in this section.

The analysis procedures for the item critique included two main tasks: 1) computing descriptive statistics (mean, standard deviation, standard error, range) for each item by using the BMD 01D program, and 2) identifying items which might have problems of ambiguity. This second analysis was performed by surrounding the mean of each item in both the clarity and ease of answering scales (see pp. 13-14) with a confidence interval. If $\bar{X} + t\sigma_{\bar{X}} > 2.5$ for an item on either of the two scales, the item was determined to be possibly ambiguous and in need of a further check. This further check was performed by referring to a notebook which had been kept on all SAES items. The notebook contained 234 pages, one for each item. Anytime a person provided a written comment about an item in addition to his numerical rating of the item, the comment was recorded in the notebook.

The mean and standard deviation for each item on the competence scale only are provided in Appendix F. In Appendix G, the items which were revised as a result of the item critique appear.

The Role Description Questionnaire critique was performed by answering five open-ended questions. Most respondents agreed that most of the questions were conducive to the checklist format. The 24 role descriptions were not particularly useful; 8 of the 11 respondents in this group had trouble identifying themselves according to the role descriptions. A point made by several respondents was that they performed several of the tasks, but for less than full time or even for less than half time. For this reason, it was planned that in the full survey, a list of roles would be provided, beside which a series of columns would be placed. Each column would represent a different amount of a person's professional time (e.g., 0-20%, 21-40%). In that format, if a person usually spent one day per week consulting on evaluation problems, he would simply place a check in the 0-20% column opposite the "Consultant:evaluation" entry.

Results of Item Contamination Study

For the competence scale only, the mean response to each double verb item was compared to the mean response of the two single verb items formed from the double verb item. For example, item 1 on the double verb list was divided into item 17 (on the list of 36 single verbs and on the list of 64 single verbs) and item 57 (on the list of 64 single verbs). A one-way analysis of variance was performed comparing the mean response of items 1, 17, and 57 on the competence scale. The purpose of having two lists of single verbs (one with 36 items and one with 64 items) was to

determine if the length of the instrument had anything to do with response rate. An analysis of variance was performed on the response rates for all three instruments (36 single verbs, 64 single verbs, 36 double verbs) and no significant difference was found.

Of the 36 double verb items, 4 had means which were significantly different when the items were divided into two single verb items (2 at the .05 level, 2 at the .025 level). The means from the eight single verb items from the original SAES were not significantly different among the three groups. From this, one could imply that the three groups were quite well matched and the four significant differences were, in fact, real differences. Since the number of items which might have been expected to differ by chance alone is fewer than two at the .05 level, the fact that four items were significantly different does seem to corroborate the belief that some information might have been lost by including two verbs in one item. It is likely, therefore, that in the SAES instrument to be used in the full survey, the double verb items would have been separated to form single verb items.

Results of Nonrespondent Bias Check

The purpose of this substudy was to see if respondents differed from nonrespondents on any variables of interest. If the two groups did differ, the results of any analyses performed on the responses would likely be biased to some unknown degree. If the groups did not differ, it was possible that the results could still be biased; but the likelihood of that occurring was considered to be less than if the groups did differ.

Chi-square tests were performed on the six variables of interest (present position title, role description, employer, length of time in

present position, highest degree, date of highest degree). No significant differences in any of these variables were found between the 15% randomly sampled nonrespondents and the 15% randomly sampled respondents. (The answers to these questions were obtained from the respondents by looking at their Role Description Questionnaire.) Thus, it may be said that at least as these variables might have affected responses to SAES items, there was little evidence of a nonrespondent bias.

In answer to the question of why the nonrespondents failed to return the SAES instrument, a total of six reasons were given. Those reasons (and the number of persons giving that reason) are shown in Table 4. Some persons gave more than one reason.

Table 4

Reasons for Nonresponse to SAES

Reason	Number of Persons Giving this Reason
Changing jobs	2
Length of the instrument	11
Not an evaluator	6
Sent in (but never received by SAES staff)	4
Too busy with other responsibilities	1
Not forwarded from previous job	2

The length of the instrument, it appears, was a very definite factor in the low response rate. Because of this, a decision was made by the SAES staff that item sampling procedures would be used in the full survey. A more detailed discussion of this topic will be given in the "Directions for Future Research" section of this report.

Results of Mailing Costs Substudy

The purpose of this substudy, it will be remembered, was to determine the most cost-effective means for improving response rates. Two forms of follow-up (postcard and letter) and three formats of return envelope (stamped, business reply, and unstamped) were used in the substudy.

Analyses of variance on the return rates of each group in the 2 x 3 factorial design were computed using the BMD 08V program. Separate analyses were performed on the home and office address groups. Analyses were performed for two time periods: 1) returns as of 12/18/72 (after one follow-up), and 2) returns as of 1/12/73 (after two follow-ups). For these analyses only, incomplete and undeliverable responses were counted as full responses.

After one follow-up, there were no significant differences in response rates among the home address group in either form of follow-up, format of return envelope, or the interaction of these two variables. Within the office address group, there was no difference among forms of follow-up or in the interaction term, but there was a difference ($p < .05$) among format of return envelope. A Tukey test showed that the business reply envelope had been returned significantly more often ($p < .05$) than the unstamped return envelope.

After the second follow-up (which, it will be remembered, was a postcard to all nonrespondents), there were again no differences among response rates for the home address group. The same pattern of significant differences occurred for the office address group as in the first analysis. This time, however, the difference among response rates for the three return envelopes was significant at the .005 level. Tukey tests showed

the business reply envelope to have been returned more often than the unstamped envelope ($p < .025$) and the stamped envelope to have been returned more often than the unstamped envelope ($p < .01$).

These results prompted the following decisions concerning the full survey to be made by the SAES staff. First, the much less expensive postcard follow-up would be used. Second, the most economical return envelope appeared to be the business reply (since payment for postage was made only on those questionnaires which were returned), and so they would be included in the full survey.

CONCLUSION

The Self Assessment of Evaluation Skills (SAES) described in this report, including the instrument's development and testing, would have been a necessary precursor to any national survey of evaluators in which the instrument was used. Similarly, instrument development activities such as those described here would be required before any kind of program planning could be carried out using the SAES instrument as a basis. The requisite first steps have been taken in the development of an instrument designed to 1) describe the members of the educational evaluation profession, 2) guide students in choosing appropriate courses of study in the evaluation field, and 3) alert program developers to areas where additional training is needed so they may incorporate those areas into evaluation curricula. Unfortunately, it appears at this time that those first steps are all that will be taken.

Pilot test results are by nature of limited generalizable value, for the intent and purpose is to precede and provide information for a larger study to follow. Therefore, these pilot test results are of limited value beyond the study for which the pilot test was to be a part. In addition to this major caution, certain other cautions should be made to persons who might read these results and wish to draw interpretations from them. These cautions are discussed as "Limitations of this Study." Similarly, some persons reading this report may see attendant areas they might like to study, and for those persons the "Directions for Future Research" section is added.

Limitations of this Study

The most serious limitation of the results reported above is the low response rate achieved in the SAES pilot test. If the response rate had been at least over 50% (and preferably in the 70% range), and if the results of the nonrespondent bias check had been the same as reported here (i.e., no significant differences between respondents and nonrespondents), the caution expressed at this time would not be so emphatic. Readers should be aware, however, that all results reported here (especially from the correlation analyses) are based on a small return rate. Although the results of the correlation analyses have been reported as preliminary evidence that there really are separate categories of skills necessary in evaluation activities, it should be remembered that these results are based on data from only 31% of the sample.

Another limitation of the study is the fact that the Role Description Questionnaire was not particularly appropriate. Respondents had trouble identifying themselves according to the role descriptors listed, and it is unclear how that identification problem affected the way in which respondents provided information contained in the questionnaire. The SAES staff believes that a more effective means for obtaining the necessary information would be the method described in the "General Results" section of this report.

As previously mentioned, it is possible that the 2,500 names contained in the SAES staff's directory of educational evaluators was over-represented by college and university personnel. A more thorough search should be conducted before a full survey is undertaken to see if particular groups have been omitted from the list of names. State departments of education, large school districts, and professional education associations (besides

AERA) should be contacted and asked for help in completing the directory.

It should also be mentioned that some of the persons to whom the questionnaire was sent wrote back to say they were not evaluators. It is likely that some proportion of the nonrespondents considered themselves not to be a member of the specified population and thus failed to return the questionnaire. (The results of the nonrespondent bias check would confirm this possibility.) The SAES staff is unclear about how this problem might be eliminated, but it is definitely a matter of concern to anyone who might be contemplating further research in this field of investigation.

The length of the SAES instrument was a definite drawback in the design of the pilot test. Item sampling procedures should have been implemented at that time; had they been, the response rate would most likely have been higher. The decision was made by the SAES staff that such procedures would be used in the full survey.

A final limitation of the study is that there is no clear evidence of the validity of the self-report data provided by respondents. With the SAES instrument as it is now, all one has is the respondent's word that he or she is highly competent in, e.g., matrix algebra. There are no performance measures included in the SAES instrument to test the validity of that statement. The inclusion of such performance measures would be a very useful addition to the SAES instrument.

Directions for Future Research

The most obvious direction for future research in this area is the previously planned full survey using the SAES instrument. Two SAES staff members have proposed such a survey and are currently attempting to obtain

funding for the study. The basic objectives of the study remain the same as when the Model Training Project was in existence, but some revisions have been made as a result of the pilot test. Specifically, the major objective of the study is as follows:

1) To provide a description of the evaluation profession

1.1) To generate a pool of items which represent the scope of competencies necessary for an evaluator to perform in various roles

1.2) To generate a pool of cognitive items which measure the competencies in subobjective 1.1

1.3) To conduct a national survey of individuals who fill evaluative positions, using the instruments designed in subobjectives 1.1 and 1.2

1.4) To provide a catalog of evaluator role profiles in terms of self-reported skills and objectively assessed cognitive behavior

Two additional objectives relate to certain methodological considerations:

1) To test the use of overlapping matrix sampling (through item sampling techniques) in a factor analytic study

2) To test the validity of self-report data by means of a comparison with objective cognitive measures

Once funding for the full survey has been obtained, a second pilot test would need to be undertaken. In this pilot test, the item sampling procedures would be tested to make sure that the data obtained in this way from respondents would be appropriate for factor analysis. Assuming the results of that pilot test were favorable, the full survey could then be conducted.

A word of caution should be injected at this point. Numerous persons have written to the MTP asking for copies of the SAES instrument. The SAES staff has been reluctant to release the instrument because of the meager reliability and validity data now available. The staff does not consider the instrument ready for use for diagnostic purposes for individuals at this time. The instrument is ready to be used to describe groups of individuals, and that is the purpose of the full survey.

In conclusion, it may be said that the Self Assessment of Evaluation Skills (SAES) instrument was developed by Model Training Project staff for several purposes. One of the purposes was to empirically validate the Universe of Evaluation Competencies, a logical grouping of skills and knowledge required to perform the complete range of evaluation activities. After numerous revisions, the SAES instrument was judged to be ready for use in a pilot test. That pilot test was conducted in late Fall, 1972. The results of the pilot test showed that with a few revisions (in items as well as in sample design), the SAES instrument was ready for use in the full survey of educational evaluators which had been planned by the Model Training Project.

The staff of the Model Training Project at The Ohio State University agreed with the following assessment of the state of training programs for educational research and research-related personnel (including evaluators:

> One of the most serious impediments to efforts to plan or conduct training programs for research or research-related personnel in education is lack of knowledge about which particular competencies or skills are most

important in conducting research and research-related activities.[1]

The completion of the research begun by the SAES pilot test and described in this report would have been a valuable contribution to the knowledge base in the area of training in educational research and evaluation. Until that knowledge base is more complete, it is unlikely that viable training programs can be maintained.

---

[1] Worthen, B. R., Anderson, R. D., & Byers, M. L. A study of selected factors related to the training of researchers, developers, diffusers, and evaluators in education. Task Force on Training Educational Research and Research-related Personnel, American Educational Research Association, November 1971. (Final report USOE Grant No. OEG-0-71-0617(520).) p. 15

APPENDIX A

Self Assessment of Evaluation Skills (SAES) Instrument

SELF ASSESSMENT OF EVALUATION SKILLS INSTRUMENT

I can . . .

1. compare and contrast instructional research and evaluation.
2. describe a recent trend of manpower supply and demand in educational evaluation.
3. design and administer evaluation studies that assess the effectiveness of projects.
4. marshal political support for evaluation activities.
5. prepare a 20-minute slide-tape presentation on problems in evaluation.
6. identify and analyze the prevalent government and foundation sources of financial support for research, development, and instruction in educational evaluation.
7. develop and implement a system for continually informing system personnel about the work of an evaluation system.
8. communicate findings from evaluation reports via television, radio, and newspapers in such a manner that the lay public can understand and critically consider the findings.
9. describe and analyze several outstanding examples of educational evaluation systems and studies.
10. design procedures which check the degree to which a curriculum package was implemented.
11. critique an evaluation report for its cost-effectiveness value.
12. develop a schedule of reporting activities.
13. design, budget, arrange for, and support internal audits of evaluation systems.
14. plan specifically for the involvement of system personnel in the development and implementation of the evaluation system.
15. define roles that must be manned in efforts to advance the science and practice of educational evaluation.
16. design and implement evaluation studies that focus on needs, problems, and opportunities within the parent agency.
17. set up a management system for evaluation programs.
18. analyze the information requirements and reading levels of various audiences for evaluation reports.
19. formulate specifications for a systems analysis.

20. communicate findings from evaluation studies in meetings in such a manner that professional educators and policy figures can use the information to influence their decisions and actions.

21. distinguish between research questions appropriate to the case study methodology and field study methodology.

22. lead policy groups to adopt a sound set of evaluation policies.

23. show that discriminant analysis is a special case of canonical correlation.

24. write both technical and popularized versions of an evaluation report, given the findings from an evaluation study.

25. write an article that is appropriate for submission to Educational Projects Information Exchange.

26. develop and implement an in-service training program in evaluation for persons at operational and managerial levels of the parent agency.

27. communicate effectively in small group settings with policy groups such as teacher associations and boards of education concerning the need for and nature of evaluation systems.

28. apply specified criteria to alternative program strategies in order to judge their relative merits.

29. describe the use of judgmental data in evaluation and techniques by which judgments should be collected.

30. develop funding proposals for programs designed to advance the profession of educational evaluation.

31. set up review procedures for articles submitted to a journal.

32. develop and/or critique a handbook of policies, organization and procedures for an evaluation system.

33. write an article that is appropriate for submission to the Journal of Educational Measurement.

34. list what audio-visual materials are available for the training of evaluators.

35. design and conduct an evaluation of an evaluation system.

36. while serving as an evaluator, develop and maintain rapport with the staff of a program being evaluated.

37. project funding strategies for the evaluation system that incorporate core internal support, external support, and a fee structure for special evaluation services.

I can . . .

38. analyze a school in order to identify evaluation questions, audiences, and information requirements.

39. convincingly address the lay public and professional educators concerning the ethical and educational merits of particular plans for collecting evaluation data.

40. develop a general description, for public consumption, of the goals, services, and overall philosophy of the evaluation system.

41. provide information which can be used to sell a curriculum package.

42. design and conduct a simulation study of alternative educational evaluation processes.

43. develop a taxonomy which can be used to categorize a body of printed matter in a specified content area.

44. discuss with policy boards of educational institutions the policy implications of evaluation.

45. formulate budgets for special evaluation studies.

46. develop and put to work an organizational structure for the evaluation system.

47. use audio-visual aids appropriately in making oral evaluation reports.

48. describe the problem of reliability and validity for criterion referenced tests.

49. describe the effect of different media on the message to be communicated.

50. conceptualize a set of performance indicators that would form a sound basis theoretically and practically for a school district's systematic evaluative information system.

51. conduct a comparative study among extant materials and a new curricular development.

52. describe the dynamics of small group behavior.

53. communicate effectively with teachers and administrators in large group settings about the meaning of and need for policies to guide evaluation practices within an educational system.

54. review and analyze the cost-effectiveness implications of developing and implementing evaluation services.

55. develop assessment strategies for the need for curricular change, given a target population.

56. articulate evaluation policies within the parent agency's policy framework.

57. present a 30-minute address on new approaches to evaluation to an audience of approximately 500.

58. present a case for the evaluation of competing instructional strategies and suggest alternate methodologies which might be used.

59. develop contracts with external agencies specifying the terms of external evaluation studies.

60. extrapolate from evaluation reports that assess competing program strategies to project cost and political implications for the competing strategies.

61. develop both long-range and short-range projections of financial requirements for the proposed evaluation system.

62. write an editorial policy statement for the operation of a refereed journal.

63. write and enter into a sound performance contract with an outside agency.

64. design specifications for evaluation reports.

65. design measuring devices which check for unintended outcomes.

66. critique a test, having read a test manual for technical information such as norm group, reliability and validity.

67. describe and analyze several major leadership programs of research, development, and instruction in evaluation.

68. establish organizational procedures for evaluation such that evaluation can have direct links to decision makers.

69. discuss the assumptions underlying equal appearing interval scaling techniques and techniques using Thurstone's Laws of Categorical Judgments.

70. organize and administer an editorial service in relation to evaluation reports.

71. make honest and systematic use of evaluation in designing and carrying through a system of public information.

72. design evaluation feedback systems that define appropriate reporting settings, content, and media, given an analysis of audiences to be served.

73. determine and provide for office space, equipment and materials needed in an evaluation system.

74. lead policy groups to adopt a sound budget for system evaluation.

75. provide stimulating leadership and direction to those who are serving in evaluation roles.

76. provide parents with various types of evaluative data for their individual child.

77. conduct survey research using a questionnaire.

78. design, budget, arrange for, and support external audits of evaluation systems.

79. negotiate a format and time schedule for evaluation reports to a client.

80. develop goals and objectives for advancing the science and practice of educational evaluation.

81. design an evaluation report to be presented by video-tape.

82. describe and analyze the major professional organizations that are active in the development of the educational evaluation profession.

83. use reference sources such as Buros' Mental Measurements Yearbook or Tests in Print for the selection of a standardized instrument.

84. design and implement a program for evaluating the operational level staff, such as teachers and school principal, of an educational agency.

85. work with program personnel to generate policies and guidelines that will govern evaluation activities in the program.

86. design and implement procedures for publishing and disseminating evaluation reports.

87. write a formal critique for publication concerning the technical adequacy of a factor analytic study of evaluation roles.

88. design charts--such as histograms, trend line graphs, cross break tables, and pie graphs--to communicate evaluation findings both to professional educators and the lay public.

89. assist in the revision of an evaluation activities schedule and in its integration in the parent agency's master schedule.

90. describe the National Assessment movement and performance contracting, drawing the implications of each for evaluation.

91. use evaluation information to design, operationalize, and implement an agency accountability system.

92. work with reference groups to anticipate decisions and associated information requirements to be served by evaluation.

93. design mechanisms to collect judgmental information concerning the worth of a curriculum package.

94. design controlled research studies of small units in a curriculum.

95. discuss the content of various articles in the AERA monograph series on curriculum evaluation.

96. develop a PERT network.

97. explain the standard procedures used by most testing companies when norming a test.

98. select, organize, and lead groups to develop criteria for judging competing program and project strategies.

99. write an article that is appropriate for submission to the Review of Educational Research.

100. describe the difference between Flanders' and Taba's Classroom Observation Schedules.

101. design and administer evaluation studies that monitor the implementation of projects.

102. give a brief outline of the content of several texts which describe evaluation models.

103. describe the evaluation system to a group of evaluation theoreticians so that they understand the basic conception of the system in relation to extant theoretical formulations in evaluation.

104. articulate current problems in evaluation and where to find reading materials about them.

105. teach teachers how to write behavioral objectives.

106. describe the use of standardized tests for placement and diagnostic purposes.

107. provide student outcome measures which can be included as part of a curriculum package.

108. recruit, orient, and maintain communication with a pool of consultants who will periodically assist in the work of the evaluation system.

109. formulate specifications for an accountability system.

110. describe basic concepts of analysis of variance, e.g., design matrix, random effect, interaction, confounding, and error.

111. describe the difference between formative and summative evaluation in terms of information needs and inferential base.

112. use evaluation information to effect rational decisions about goals, strategies, management, and the recycling of activities.

113. involve the lay public and professional educators in determining the questions to be addressed by an evaluation system.

114. Recruit, select, and orient core staff for an evaluation system.

115. communicate clearly and convincingly over television and radio about the nature and importance of an evaluation system.

116. critique an evaluation report for its relevance to real decision problems.

I can . . .

117. construct simulation exercises which portray real evaluative problems.

118. develop and disseminate clear and concise descriptions of evaluation systems that can be understood by the public.

119. write forced-choice and free-choice test items.

120. organize groups of influential and professional personnel for performing tasks related to improving the evaluation profession.

121. define and maintain up-to-date role descriptions and salary schedules for evaluation personnel.

122. design and carry through an analysis of variance using a hand calculator.

123. distinguish between revision cycles for Thurstonian Attitude Scales and Likert Scales.

124. write an analysis of the relevance of economic theory to educational evaluation.

125. design instruments which measure the cognitive behaviors delineated in a curriculum package.

126. understand and judge the potential utility of outputs from projected data analyses.

127. define a population and formulate sampling specifications for data collection purposes.

128. describe different usage of COBOL, FORTRAN, and assembly language to select the language best suited to a given evaluation problem.

129. understand basic concepts of matrix algebra as they relate to statistics, e.g., inversion, transposition, rank, characteristic equation, conformability, trace.

130. anticipate decisions to be served by special evaluation studies.

131. critique an evaluation design in terms of analysis and measurement assumptions.

132. develop and present a major address on the state of the art in educational evaluation to audiences of educational researchers.

133. describe the application of and the relationships existing among the major theoretical distributions of statistics, e.g., the normal, binomial, Poisson, chi-square, variance-ratio, and Cauchy distributions.

134. generate criteria with potential applicability for evaluating some generally defined developmental program such as "Individually Prescribed Instruction."

135. discuss various definitions of accountability and distinguish among accountability, evaluation, and assessment.

136. present a case for the evaluation of program goals.

137. identify alternative analysis techniques and compare their appropriateness for a given set of data specifications which spell out design, sampling, and measurement assumptions.

138. read with understanding most articles in Psychometrika.

139. design an experiment so as to allow analysis of interactions between treatment and classification factors.

140. design and implement a multiple regression type of study to assess the influence of several variables, including provision of evaluation information on decision-making behavior.

141. describe the basic structure and constructs of the FORTRAN language, e.g., type, dimension, function, equivalence, format.

142. describe how validity coeffecients are calculated and what they mean.

143. compare and contrast the definition of true score in traditional measurement theory and the definition of universe score in generalizability theory.

144. perform an item analysis and interpret the results.

145. write a formal critique for publication concerning the technical adequacy of a field experiment comparing alternative evaluation techniques.

146. discuss the difference between orthogonal comparison and trend analysis.

147. distinguish between a psychometric inference and a statistical inference.

148. project and define a number of research problems that might be pursued programmatically.

149. develop and implement means for responding quickly and directly to requests for assistance in designing special evaluation studies.

150. design an evaluation of extant curricular packages in a specified content area.

151. describe basic concepts involved in parameter estimation, e.g., bias, consistency, likelihood ratio, etc.

152. write operational objectives and specify attainment criteria associated with them.

153. describe the functions of an evaluator in a large curriculum development project.

154. design and develop a data bank.

155. identify and select from among the available measuring instruments, given a set of information requirements.

156. develop forms and procedures for managing the data processing operations of an evaluation system.

I can . . .

157. discuss critically emergent developments in the field of educational measurement.

158. interpret test results given in standard score format such as grade equivalent, age equivalent, percentiles, deciles, or stanines.

159. write a test blueprint.

160. design procedures for assessing various media presentations of curriculum content.

161. suggest techniques appropriate to measure degree of attainment for affective objectives.

162. write an analysis of the relevance of demographic analysis to educational evaluation.

163. design a sampling plan for field testing materials.

164. understand evaluation reports which incorporate discriminant analysis, factor analysis, regression analysis, and multivariate analysis of variance and covariance.

165. organize and administer a standardized testing program.

166. design and implement a field experiment comparing alternative evaluation techniques, e.g., item sampling versus examinee sampling.

167. analyze data from free response interview data collection methods.

168. design and conduct a historical study of the development of educational evaluation theory and practice covering a 10-year period.

169. distinguish between situations where erosive unobtrusive measures are more appropriate than accretive measures.

170. discuss the logic of statistical analysis, the major classes of questions that can be addressed by present modes of analysis, the classes of assumptions that can be accommodated by present statistical technology, and the emergent developments in the field.

171. apply appropriately the major nonparametric techniques, e.g., Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman, Kolmogorov-Smirnoff.

172. identify, compare, and contrast the major approaches to measurement that have application in educational evaluation.

173. formulate a researchable hypothesis appropriate to a tightly controlled laboratory study.

174. describe basic theoretical concepts relating to distribution-free statistics, e.g., robustness, asymptotic relative efficiency, stochastic inequality, location and scale.

175. provide strategies to assess content for inclusion in or exclusion from proposed curriculum packages.

176. select a test that has been normed on an appropriate group.

177. understand evaluation reports which incorporate means, standard deviations, percentiles, grade-equivalent norms, correlations, and analyses of variance.

178. describe evaluation functions served by various offices within a university.

179. compare the predictive validity of performance measures and simulation exercises.

180. identify and define briefly major theories relating to learning and development, e.g., Hull, Piaget, Gagné, Skinner, Bruner.

181. discuss the behavioral objectives controversy in evaluation enumerating the relationship between the application of each evaluation model and the model's dependence on behavioral objectives.

182. review major writings in the area of decision-making theory and incorporate their main ideas into a discussion of the utility of extant theoretical work in educational evaluation.

183. systematically assess and judge the merits of an evaluation plan.

184. develop a monograph that identifies, compares, and contrasts the major extant theoretical formulations in the field of educational evaluation.

185. design and carry through computer-based analysis of variance, multiple regression studies, and factor analysis using canned programs.

186. describe and analyze the present state of development of the educational evaluation profession, as contrasted with those of educational research and development, educational administration, and psychology.

187. design and administer programs of research, development, instruction, and service for the improvement of evaluation.

188. describe the historical development of evaluation models and compare and contrast focal points of each.

189. conduct a literature review, write it up, and present it to a curriculum development group, irrespective of substantive area.

190. write affective objectives at various levels of the affective taxonomy.

191. relate the major research strategies and substantive areas of philosophy and history to evaluation problems.

192. design measuring instruments for affective objectives of a curriculum package.

193. formulate and explain a set of criteria for judging theoretical work in educational evaluation.

194. interpret standardized test results to parents and students.

195. present the main principles of information theory and use them to critique the adequacy of present models of educational evaluation.

196. state how reliability coefficients are calculated and what they mean.

197. chair a curriculum evaluation committee.

198. compare and contrast the roles of an internal and external evaluator in terms of focus, information sources, and credibility.

199. develop budget forms and procedures for an evaluation system.

200. delineate evaluation authority and responsibility within the agency's organizational structure.

201. organize and administer a data processing center related to meeting evaluation objectives.

202. compare and contrast traditional site visit methodology to advocate team methodology.

203. list and describe major information processing concepts, e.g., multiprogramming, random access, buffered I/O, indirect and relative addressing, and subroutining.

204. design a study to evaluate the goals of a new curricular development.

205. critique an experimental design in terms of its internal and external validity.

206. compare and contrast the nature of educational evaluation with such related fields as research, development, planning, management.

207. relate the major research strategies and theories of sociology, political science and economics to evaluation problems.

208. set up a framework within which program personnel can generate testable creative solution strategies for specified problems.

209. interpret the computer printout for discriminant function analysis and MANOVA studies.

210. provide specifications for and can control a management information system that will maintain up-to-date information about program and project events and activities.

211. select, organize, and lead groups of professionals to generate alternative program strategies and project designs.

212. design a market research study for a new product.

213. establish and implement quality control procedures for an information system.

214. relate the major research strategies and substantive areas of experimental and social psychology to evaluation problems.

I can . . .

215. use the Delphi technique to assist groups to identify and assign priorities to needs, problems, and opportunities.

216. evaluate and/or design and conduct a case study of an operational evaluation system.

217. list and describe the characteristics and capabilities of major information storage media, e.g., punched cards, magnetic tape, rum and disk.

218. design and implement (including proper orientation for teachers) the conditions under which data are to be gathered.

219. analyze the data produced in classroom observational study.

220. suggest techniques appropriate to evaluate objectives at various levels of the cognitive taxonomy.

221. describe evaluation theories, models, and practices in a wide range of fields outside of education.

222. design a study where path analysis is the most appropriate analysis technique.

223. design procedures for assessing competing sequences of curricular content.

224. design an information system according to which data are to be coded, stored, and retrieved.

225. specify, operationalize, and apply criteria for evaluating evaluation systems.

226. organize and administer a public information service in relation to evaluation information.

227. defend the choice of an oblique rotation or an orthogonal rotation in a factor analysis study.

228. state the basic principles of value theory and utility theory and compare and contrast the relevance of these fields for theory development in educational evaluation.

229. write objectives at each level of the cognitive taxonomy.

230. write a criterion referenced test.

231. review historically the development of the legal bases for evaluation in education.

232. conduct a training session for participant observers.

233. design and implement evaluation studies that identify and assess competing solution strategies for specified problem areas.

234. describe and analyze user attitudes toward evaluation.

APPENDIX B

Attachments to SAES in Pilot Test

Commission for the Study of the Evaluation Profession

Evaluation Center / College of Education / The Ohio State University / Columbus, Ohio 43210

November 10, 1972

Ms. Anne P. Taylor
P. O. Box 603
Corrales, New Mexico 87408

Dear Ms. Taylor:

With support from the U.S. Office of Education, the Evaluation Center of The Ohio State University, in conjunction with the Commission for the Study of the Evaluation Profession, is collecting data to determine the range of competencies and interests in evaluation skills possessed by practicing evaluators, and the relative importance of these skills in job performance. A full survey of the evaluation profession is planned for early Spring; at present, we are concerned with refining the instruments to be used to collect these data. You are receiving this letter because your name was drawn as part of a nationwide random sample of persons interested in educational evaluation.

Several instruments are enclosed with this letter. One is the Self Assessment of Evaluation Skills (SAES). The items in this instrument comprise a sample of skills chosen from the Universe of Evaluatinn Competencies. The Universe was developed through research at Ohio State combined with results from previous studies of evaluation skill requirements. In the full survey of the evaluation profession, the task will be to determine how these skills relate to different evaluation roles. The data to be collected will be used to describe the skill configurations in various norm groups. The Role Description Questionnaire is included so we may gather information to be used to determine those norm groups.

We would like you to complete these instruments as if you were a participant in the full survey so that we may do preliminary analyses to see if the instruments are functioning as intended. In addition, the pilot sample has been divided into six subsamples; each subsample is being asked to critique the technical adequacy of parts of the instruments. Full directions for performing these critiques appear later. All of

your answers will be completely confidential; only group statistics will be reported. The answer sheets and other materials are numbered to facilitate checking off responses as they are received.

We realize that this is a big task we are asking you to perform. We are convinced, however, that this research is necessary to the growth of the field of educational evaluation. Evaluation training programs need to train students so they will be qualified to accept "real world" positions in evaluation; but to date so little research has been done in this area that training requirements relevant to real evaluation practice are unknown. Your participation in this study will add to the knowledge base in the field of educational evaluation, and your cooperation will be greatly appreciated.

Please return the SAES instrument, Role Description Questionnaire, and all answer sheets in the pre-addressed return envelope which has been provided for your convenience. If you would like to have your name placed on the mailing list for the Interfacer, the newsletter of the Evaluation Center's Model Training Project, please fill in the yellow card asking for your name and address and enclose it with the other materials in the return envelope.

Thank you very much for your help in this endeavor.

Sincerely,

Mary Anne Bunda

Daniel L. Stufflebeam

Daniel L. Stufflebeam
Study Co-directors

Enclosures

Directions for Completing the Self Assessment of Evaluation Skills (SAES)

The purpose of this study is to collect data on an evaluator's competence and interest in specific areas of evaluation and related activities and his perception of the relative importance of each of these areas to his professional field. These norm group data, collected from a wide range of people presently engaged in evaluation activities, will be used to validate the instrument so that it can be used to establish training programs for students in evaluation and to aid in the placement of students in various agencies. Complete data is of the utmost importance. Please answer each item thoroughly and honestly. Your response will be held in strictest confidence; all data will be reported by group statistics only.

Each item will be rated according to three different scales: Competence, Interest, and Importance. Read each of the following scales carefully, then look at the examples. These examples illustrate how the three scales can be combined to form a single point of view concerning a particular item. After studying the scales and examples, mark each item according to the three frames of reference by filling in the blank space under each scale with the number from that scale which best describes your rating of that item. Be sure to fill in all three scales for each item.

Competence Scale

1. No Competence - Respondent has no base from which to respond and would be completely unable to perform this task.

2. Minimum Competence - Respondent could perform this task minimally well only with extensive study. He does, however, know enough about the area to hire a knowledgeable consultant.

3. Moderate Competence - Respondent could perform this task moderately well with some study and well with extensive study.

4. High Competence - Respondent could perform this task well with minimal study and exceptionally well with extensive study.

5. Superior Competence - Respondent could perform this task exceptionally well with little or no study.

Interest Scale

1. Negative Interest - Respondent finds this area repugnant.

2. No Interest - Respondent has absolutely no interest in this area.

3. Minimal Interest - Respondent finds the area somewhat interesting but can generally find more desirable activities to engage in.

4. Moderate Interest - Respondent has genuine interest in this area but there are professional activities he finds more desirable.

5. High Interest - There are few professional activities that the respondent would rather engage in.

Importance Scale

1. No Importance - Respondent almost never performs this task and considers it irrelevant to his professional performance.

2. Minimal Importance - Respondent occasionally performs this task but considers it irrelevant to his professional performance.

3. Moderate Importance - Respondent occasionally performs this task and considers it relevant to his professional performance.

4. High Importance - Regardless of the frequency of performance of this task, respondent considers it relevant to his professional performance.

5. Crucial - Ability to perform this task, regardless of its frequency, is vital to the respondent's successful functioning in his profession.

Examples

Please refer to the preceding scales:

Item: Can provide for necessary equipment acquisition and maintenance.

Competence	Interest	Importance
(4)	(2)	(5)

The respondent feels that he is highly competent in this activity but he has no interest in it, even though he thinks the area is a crucial one to his evaluation unit.

Item: Can write a questionnaire.

Competence	Interest	Importance
(4)	(5)	(2)

The respondent feels that he is highly competent in this area and he finds it highly interesting, although he thinks it is irrelevant to his particular professional field.

Item: Can establish criteria for evaluating an evaluation system.

Competence	Interest	Importance
(2)	(4)	(4)

Even though the respondent feels he is only minimally competent in this area, he has a genuine interest in the area and feels it is of high importance.

The results of each individual instrument will be confidential. Only the combined norm group data will be disseminated. Please feel free to answer each item frankly. Please do not omit any of the items.

Thank you for your cooperation.

ANSWER SHEET FOR SELF ASSESSMENT OF EVALUATION SKILLS (SAES)

DIRECTIONS: In each set of parentheses below, write the number which corresponds best to your perceived degree of competence, interest, and importance for each of the items in the Self Assessment of Evaluation Skills (SAES). Write one number for each of the three scales for each item. Shortened definitions for each point on the scales are given on each page of this answer sheet; for a more detailed description, please refer to the "Directions for Completing the Self Assessment of Evaluation Skills" which accompany these materials.

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest - area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest - area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevant to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP
1.	( )	( )	( )	7.	( )	( )	( )	13.	( )	( )	( )	19.	( )	( )	( )
2.	( )	( )	( )	8.	( )	( )	( )	14.	( )	( )	( )	20.	( )	( )	( )
3.	( )	( )	( )	9.	( )	( )	( )	15.	( )	( )	( )	21.	( )	( )	( )
4.	( )	( )	( )	10.	( )	( )	( )	16.	( )	( )	( )	22.	( )	( )	( )
5.	( )	( )	( )	11.	( )	( )	( )	17.	( )	( )	( )	23.	( )	( )	( )
6.	( )	( )	( )	12.	( )	( )	( )	18.	( )	( )	( )	24.	( )	( )	( )
	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest - area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest - area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevant to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

---

	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP
25.	( )	( )	( )	35.	( )	( )	( )	45.	( )	( )	( )	55.	( )	( )	( )
26.	( )	( )	( )	36.	( )	( )	( )	46.	( )	( )	( )	56.	( )	( )	( )
27.	( )	( )	( )	37.	( )	( )	( )	47.	( )	( )	( )	57.	( )	( )	( )
28.	( )	( )	( )	38.	( )	( )	( )	48.	( )	( )	( )	58.	( )	( )	( )
29.	( )	( )	( )	39.	( )	( )	( )	49.	( )	( )	( )	59.	( )	( )	( )
30.	( )	( )	( )	40.	( )	( )	( )	50.	( )	( )	( )	60.	( )	( )	( )
31.	( )	( )	( )	41.	( )	( )	( )	51.	( )	( )	( )	61.	( )	( )	( )
32.	( )	( )	( )	42.	( )	( )	( )	52.	( )	( )	( )	62.	( )	( )	( )
33.	( )	( )	( )	43.	( )	( )	( )	53.	( )	( )	( )	63.	( )	( )	( )
34.	( )	( )	( )	44.	( )	( )	( )	54.	( )	( )	( )	64.	( )	( )	( )
	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest - area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest - area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevant to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP
65.	( )	( )	( )	75.	( )	( )	( )	85.	( )	( )	( )	95.	( )	( )	( )
66.	( )	( )	( )	76.	( )	( )	( )	86.	( )	( )	( )	96.	( )	( )	( )
67.	( )	( )	( )	77.	( )	( )	( )	87.	( )	( )	( )	97.	( )	( )	( )
68.	( )	( )	( )	78.	( )	( )	( )	88.	( )	( )	( )	98.	( )	( )	( )
69.	( )	( )	( )	79.	( )	( )	( )	89.	( )	( )	( )	99.	( )	( )	( )
70.	( )	( )	( )	80.	( )	( )	( )	90.	( )	( )	( )	100.	( )	( )	( )
71.	( )	( )	( )	81.	( )	( )	( )	91.	( )	( )	( )	101.	( )	( )	( )
72.	( )	( )	( )	82.	( )	( )	( )	92.	( )	( )	( )	102.	( )	( )	( )
73.	( )	( )	( )	83.	( )	( )	( )	93.	( )	( )	( )	103.	( )	( )	( )
74.	( )	( )	( )	84.	( )	( )	( )	94.	( )	( )	( )	104.	( )	( )	( )
	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP

55

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest - area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest - area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevant to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

---

	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP
105.	( )	( )	( )	115.	( )	( )	( )	125.	( )	( )	( )	135.	( )	( )	( )
106.	( )	( )	( )	116.	( )	( )	( )	126.	( )	( )	( )	136.	( )	( )	( )
107.	( )	( )	( )	117.	( )	( )	( )	127.	( )	( )	( )	137.	( )	( )	( )
108.	( )	( )	( )	118.	( )	( )	( )	128.	( )	( )	( )	138.	( )	( )	( )
109.	( )	( )	( )	119.	( )	( )	( )	129.	( )	( )	( )	139.	( )	( )	( )
110.	( )	( )	( )	120.	( )	( )	( )	130.	( )	( )	( )	140.	( )	( )	( )
111.	( )	( )	( )	121.	( )	( )	( )	131.	( )	( )	( )	141.	( )	( )	( )
112.	( )	( )	( )	122.	( )	( )	( )	132.	( )	( )	( )	142.	( )	( )	( )
113.	( )	( )	( )	123.	( )	( )	( )	133.	( )	( )	( )	143.	( )	( )	( )
114.	( )	( )	( )	124.	( )	( )	( )	134.	( )	( )	( )	144.	( )	( )	( )
	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest - area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest - area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevant to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP
145.	( )	( )	( )	155.	( )	( )	( )	165.	( )	( )	( )	175.	( )	( )	( )
146.	( )	( )	( )	156.	( )	( )	( )	166.	( )	( )	( )	176.	( )	( )	( )
147.	( )	( )	( )	157.	( )	( )	( )	167.	( )	( )	( )	177.	( )	( )	( )
148.	( )	( )	( )	158.	( )	( )	( )	168.	( )	( )	( )	178.	( )	( )	( )
149.	( )	( )	( )	159.	( )	( )	( )	169.	( )	( )	( )	179.	( )	( )	( )
150.	( )	( )	( )	160.	( )	( )	( )	170.	( )	( )	( )	180.	( )	( )	( )
151.	( )	( )	( )	161.	( )	( )	( )	171.	( )	( )	( )	181.	( )	( )	( )
152.	( )	( )	( )	162.	( )	( )	( )	172.	( )	( )	( )	182.	( )	( )	( )
153.	( )	( )	( )	163.	( )	( )	( )	173.	( )	( )	( )	183.	( )	( )	( )
154.	( )	( )	(( )	164.	( )	( )	( )	174.	( )	( )	( )	184.	( )	( )	( )
	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest - area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest - area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevant to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP
185.	( )	( )	( )	195.	( )	( )	( )	205.	( )	( )	( )	215.	( )	( )	( )
186.	( )	( )	( )	196.	( )	( )	( )	206.	( )	( )	( )	216.	( )	( )	( )
187.	( )	( )	( )	197.	( )	( )	( )	207.	( )	( )	( )	217.	( )	( )	( )
188.	( )	( )	( )	198.	( )	( )	( )	208.	( )	( )	( )	218.	( )	( )	( )
189.	( )	( )	( )	199.	( )	( )	( )	209.	( )	( )	( )	219.	( )	( )	( )
190.	( )	( )	( )	200.	( )	( )	( )	210.	( )	( )	( )	220.	( )	( )	( )
191.	( )	( )	( )	201.	( )	( )	( )	211.	( )	( )	( )	221.	( )	( )	( )
192.	( )	( )	( )	202.	( )	( )	( )	212.	( )	( )	( )	222.	( )	( )	( )
193.	( )	( )	( )	203.	( )	( )	( )	213.	( )	( )	( )	223.	( )	( )	( )
194.	( )	( )	( )	204.	( )	( )	( )	214.	( )	( )	( )	224.	( )	( )	( )
	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest - area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest - area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevant to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

---

	COMP	INT	IMP
225.	( )	( )	( )
226.	( )	( )	( )
227.	( )	( )	( )
	COMP	INT	IMP

	COMP	INT	IMP
228.	( )	( )	( )
229.	( )	( )	( )
230.	( )	( )	( )
	COMP	INT	IMP

	COMP	INT	IMP
231.	( )	( )	( )
232.	( )	( )	( )
	COMP	INT	IMP

	COMP	INT	IMP
233.	( )	( )	( )
234.	( )	( )	( )
	COMP	INT	IMP

Role Description Questionnaire

The following items will provide us with descriptive data concerning your present position and previous position. Please fill in each blank. A list of roles is attached to which you may refer. Feel free to indicate combinations of the roles which are described below or add new roles. If you use one of the roles listed, please list the name of the role and any additional information specified in that description.

Present Position

Title_____

Role Description/Comments (see list below)_____

_____

Department or Unit_____

Institution or Agency_____

Length of time at present position as of Oct. 1, 1972_____

Degree required for position_____

Certification required for position_____

Most Recent Previous Position

Title_____

Role Description/Comments (see list below)_____

_____

Department or Unit_____

Institution or Agency_____

Length of time at previous position_____

Degree required for position_____

Certification required for position_____

Role Description Questionnaire (Cont.)

Degree Held - please supply information for the two most recently obtained

	Most Recent	Next Most Recent
Degree Held		
College or University		
Year Completed		
Major Field		
Specific Area		

Age _____ Sex ___ M ___ F

Role Descriptions

1. Administrator of Research - majority of your time is spent directing research projects.

2. Administrator of Evaluation - majority of your time is spent directing evaluation projects (reviewing designs, budgets, proposals).

3. Other Academic Administrator - majority of time is spent administering in an area other than research or evaluation.

4. Teacher of Research - most time spent teaching research related courses (design, statistics, measurement). Please specify course area.

5. Teacher of Evaluation - most time spent teaching evaluation courses (evaluation theory, design). Please specify course area.

6. Teacher of Development - majority of time is spent teaching development related courses (techniques of packaged courses, innovations in education). Please specify course area.

7. Other University Instructor - majority of time is spent as an instructor in an area other than RDD and E. Please specify area.

8. Elementary / Secondary Teacher - majority of time is spent as a teacher in elementary or secondary schools.

9. 50% Teacher / 50% Researcher - time is split equally between teaching and research. Please note fields in which you teach and do research.

Role Descriptions (Cont.)

10. 50% Teacher / 50% Administrator - time is split equally between teaching and administrating. Please note field in which you teach.

11. 50% Researcher / 50% Administrator - time is split equally between research and administration. Please note field in which you do research.

12. 50% Teacher / 50% Evaluator - time is split equally between teaching and evaluating. Please note field in which you teach.

13. 50% Administrator / 50% Evaluator - time is split equally between administrating and evaluating.

14. 50% Researcher / 50% Evaluator - time is split equally between research and evaluation. Please note field in which you do research.

15. Product Developer - majority of time spent developing educational products (packaged courses, texts).

16. Diffusion Specialist - majority of time is spent in diffusion efforts (report writing, and presenting).

17. Evaluation Specialist; Design - majority of time is spent developing designs for evaluation of educational efforts.

18. Data Analyst - most of time is spent analyzing data collected by others.

19. Statistician - most of time is spent consulting with projects on statistical questions.

20. Counseling - most of time is spent counseling students in school systems or universities.

21. Director of Standardized Testing - most of time spent directing efforts of school system or university, in collecting standardized test data.

22. Evaluation Specialist; Implementation - majority of time is spent in the implementation of an evaluation design (e.g. data collection, reporting).

23. Federal Project Evaluator - majority of time is spent in the evaluation of a federally funded program or project. Please specify type of project.

24. Evaluation Consultant - majority of time is spent in some other endeavor but often consultant help provided to other agencies in evaluation.

Other - please list and define _____

_____

_____

Directions and Answer Sheet for Critiquing Items

Besides being interested in your answers to the items on the three scales, we would appreciate it very much if you would help us assess the technical adequacy of the SAES instrument. We would like you to pay specific attention to 50 items, item numbers 1 through 50. (Other members of the sample are critiquing the remaining 184 items.) Please rate the clarity and ease of understanding of each of the 50 items by marking the appropriate option in the questions below. If you have written comments about the items, we would be very interested in seeing them.

1. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

2. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

3. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

4. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

5. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

6. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

7. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

8. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

9. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

10. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

11. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

12.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

13.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

14.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

15.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

16.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

17.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

18.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

19.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

20.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

21.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

22.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

23.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

24.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

25.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

26.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

27.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

28.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

29.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

30.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

31.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

32.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

33. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

34. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

35. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

36. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

37. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

38. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

39. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

40.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

41.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

42.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

43.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

44.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

45.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

46.

I found this item	I found this item
( ) completely understandable	( ) difficult to answer
( ) moderately clear	( ) moderately difficult
( ) moderately ambiguous	( ) moderately easy
( ) completely ambiguous	( ) easy to answer

Comments_____

47. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

48. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

49. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

50. I found this item
( ) completely understandable
( ) moderately clear
( ) moderately ambiguous
( ) completely ambiguous

I found this item
( ) difficult to answer
( ) moderately difficult
( ) moderately easy
( ) easy to answer

Comments_____

Directions and Answer Sheet for Critiquing Role Description Questionnaire

Besides being interested in your answers to the items on the Role Description Questionnaire, we would appreciate it very much if you would help us assess the technical adequacy of the questionnaire. It should be noted that this is a preliminary version of the questionnaire; in the full survey, we plan to have checklists for the answers to many of the questions. We would appreciate your comments on the item below.

1. What sections of the questionnaire would be most conducive to the checklist format (e.g., for "Degree Held," a list of degrees would be provided; the respondent would check the degree which he holds)? _____

2. Is the information requested in the "Present Position" and "Most Recent Previous Position" sections clear? Do you have suggestions for clarifying the item if you feel it is necessary? _____

3. Is the information requested in the "Degree Held" section clear? Do you have suggestions for clarifying the item if you feel it is necessary? _____

4. Are the role descriptions provided clear? Do you feel they are sufficient? Are there evaluation roles which were omitted and which you feel should be listed?

5. Did you have trouble identifying yourself according to the role descriptions? If you did, how might the problem be alleviated?

APPENDIX C

Follow-up Correspondence

model training project

a new direction in education

Commission for the Study of the Evaluation Profession

Evaluation Center / College of Education / The Ohio State University / Columbus, Ohio 43210

November 30, 1972

Dear Colleague:

Recently you received a letter from us requesting your assistance in a pilot test of the Self Assessment of Evaluation Skills instrument. To date, we have not received your completed questionnaire. Your response is most important in order to have an adequate critique of the items in the instrument.

In case you might have misplaced the previous mailing, we are enclosing copies of the materials we sent to you earlier. You will find another return envelope enclosed for your convenience. Your cooperation in completing and returning the materials will be greatly appreciated.

Sincerely,

Mary Anne Bunda

Daniel L. Stufflebeam

Daniel L. Stufflebeam
Study Co-directors

model training project

new direction in education

Commission for the Study of the Evaluation Profession

Evaluation Center / College of Education / The Ohio State University / Columbus, Ohio 43210

December 15, 1972

Dear Colleague:

This is a last ditch effort -- there will be no further follow-up harassment beyond this reminder. But please, if it is at all possible, would you complete and return the Self Assessment of Evaluation Skills (SAES) instrument which we sent to you in November.

In case you might have misplaced the previous mailing, we are enclosing copies of the materials we sent to you earlier. You will find another return envelope enclosed for your convenience. Your cooperation will be most appreciated.

Sincerely,

Mary Anne Bunda

Daniel L. Stufflebeam

Study Co-directors

November 30, 1972

Dear Colleague:

Recently you received a letter from us requesting your assistance in a pilot test of the Self Assessment of Evaluation Skills instrument. To date, we have not received your completed questionnaire. Your response is most important in order to have an adequate critique of the items in the instrument. So please, if you have not already done so, return the completed questionnaire in the envelope which was provided for you. Your cooperation will be greatly appreciated.

Sincerely,

Mary Anne Bunda Daniel L. Stufflebeam

Study Co-directors

December 15, 1972

Dear Colleague

This is a last ditch effort -- there will be no further follow-up harassment beyond this reminder. But please, if it is at all possible, would you complete and return the Self Assessment of Evaluation Skills (SAES) instrument which we sent to you in November. Your cooperation will be most appreciated.

Sincerely,

Mary Anne Bunda Daniel L. Stufflebeam

Study Co-directors

APPENDIX D

Item Contamination Substudy

odel training project

new direction in education

Commission for the Study of the Evaluation Profession

Evaluation Center / College of Education / The Ohio State University / Columbus, Ohio 43210

November 17, 1972

Mr. Anthony H. Goeree
Center for Vocational & Technical Education
1900 Kenny Road
Columbus, Ohio 43210

Dear Mr. Goeree:

With support from the U.S. Office of Education, the Evaluation Center of The Ohio State University, in conjunction with the Commission for the Study of the Evaluation Profession, is collecting data to determine the range of competencies and interests in evaluation skills possessed by practicing evaluators, and the relative importance of these skills in job performance. A full survey of the evaluation profession is planned for early Spring; at present, we are concerned with refining the instruments to be used to collect these data. You are receiving this letter because your name was drawn as part of a nationwide random sample of persons interested in educational evaluation.

Two instruments are enclosed with this letter. One is SAES-S, a shortened version of the Self Assessment of Evaluation Skills. The items in the full SAES instrument comprise a sample of skills chosen from the Universe of Evaluation Competencies. The Universe was developed through research at Ohio State combined with results from previous studies of evaluation skill requirements. In the full survey of the evaluation profession, the task will be to determine how these skills relate to different evaluation roles. The data to be collected will be used to describe the skill configurations in various norm groups. The Role Description Questionnaire is included so we may gather information to be used to determine those norm groups.

We would like you to complete these instruments as if you were a participant in the full survey so that we may do preliminary analyses to see if the instruments are functioning as intended. All of your answers will be completely confidential; only group statistics will be reported. The answer sheets and other materials are numbered to facilitate checking off responses as they are received.

We feel that this research is necessary to the growth of the field of educational evaluation. Evaluation training programs need to train students so they will be qualified to accept "real world" positions in evaluation, but to date so little research has been done in this area that training requirements relevant to real evaluation practice are unknown. Your participation in this study will add to the knowledge base in the field of educational evaluation, and your cooperation will be greatly appreciated.

Please return the answer sheets in the pre-addressed return envelope which has been provided for your convenience. If you would like to have your name placed on the mailing list for the Interfacer, the newsletter of the Evaluation Center's Model Training Project, please fill in the form asking for your name and address and enclose it with the other materials in the return envelope.

Thank you very much for your help in this endeavor.

Sincerely,

Mary Anne Bunda

Daniel L. Stufflebeam

Mary Anne Bunda
Daniel L. Stufflebeam
Study Co-directors

Enclosures

I can . . .

1. present a case for the evaluation of competing instructional strategies and suggest alternate methodologies which might be used for the evaluation of these strategies.

   57. present a case for the evaluation of competing instructional strategies.

   17. suggest alternate methodologies which might be used for the evaluation of competing instructional strategies.

2. design and implement procedures for publishing and disseminating evaluation reports.

   62. design procedures for publishing and disseminating evaluation reports.

   23. implement previously designed procedures for publishing and disseminating evaluation reports.

3. design and conduct an evaluation of an evaluation system.

   55. design an evaluation of an evaluation system.

   63. conduct an evaluation of an evaluation system given an appropriate design.

4. design and implement evaluation studies that identify and assess competing solution strategies for specified problem areas.

   16. design evaluation studies that identify and assess competing solution strategies for specified problem areas.

   60. implement, given an appropriate design, evaluation studies that identify and assess competing solution strategies for specified problem areas.

5. describe the use of judgmental data in evaluation and techniques by which judgments should be collected.

   38. describe the use of judgmental data in evaluation.

   28. describe techniques by which judgments should be collected.

6. design, budget, arrange for, and support external audits of evaluation systems.

   43. design external audits of evaluation systems.

   8. budget, arrange for, and support external audits of evaluation systems.

7. develop and put to work an organizational structure for the evaluation system.

   25. Implement a previously developed organizational structure for the evaluation system.

I can . . .

21. develop an organizational structure for the evaluation system.

9. develop and implement a system for continually informing parent agency personnel about the work of an evaluation unit.

12. develop a system for continually informing parent agency personnel about the work of an evaluation unit.

31. implement a previously developed system for continually informing parent agency personnel about the work of an evaluation unit.

10. design and develop a data bank.

16. design a data bank.

59. develop a data bank, given an appropriate design.

11. develop and implement an in-service training program in evaluation for persons at all levels of the parent agency.

20. implement an established in-service training program in evaluation for persons at all levels of the parent agency.

29. develop an in-service training program in evaluation for persons at all levels of the parent agency.

13. design and implement evaluation studies that focus on needs, problems, and opportunities within the parent agency.

39. design evaluation studies that focus on needs, problems, and opportunities within the parent agency.

3. implement designed evaluation studies that focus on needs, problems, and opportunities within the parent agency.

15. design and implement a sound program of staff evaluation.

22. design a sound program of staff evaluation.

58. implement, given an appropriate design, a sound program of staff evaluation.

16. design and implement a field experiment comparing alternative evaluation techniques, e.g., item sampling versus examinee sampling.

54. design a field experiment comparing alternative evaluation techniques, e.g., item sampling versus examinee sampling.

27. implement, given a design, a field experiment comparing alternative evaluation techniques, e.g., item sampling versus examinee sampling.

17. design and administer evaluation studies that assess the effectiveness of projects.

7. design evaluation studies that assess the effectiveness of projects.

64. administer, given an appropriate design, evaluation studies that assess the effectiveness of projects.

19. provide specifications for and control a management information system that will maintain up-to-date information about program and project events and activities.

5. provide specifications for a management information system that will maintain up-to-date information about program and project events and activities.

14. control, given specifications, a management information system that will maintain up-to-date information about program and project events and activities.

20. use evaluation information to design and implement an agency accountability system.

36. use evaluation information to design an agency accountability system.

40. implement, given an appropriate design, an agency accountability system.

22. design and implement quality control procedures for an information system.

41. design quality control procedures for an information system.

56. implement previously designed quality control procedures for an information system.

23. organize and administer a data processing center related to meeting evaluation objectives.

34. organize a data processing center related to meeting evaluation objectives.

52. administer a previously organized data processing center related to meeting evaluation objectives.

24. organize and administer an editorial service in relation to evaluation reports.

2. organize an editorial service in relation to evaluation reports.

48. administer a previously organized editorial service in relation to evaluation reports.

26. design and conduct a historical study of the development of educational evaluation theory and practice covering a 10-year period.

I can . . .

50. design a historical study of the development of educational evaluation theory and practice covering a 10-year period.

47. conduct, given an appropriate design, a historical study of the development of educational evaluation theory and practice covering a 10-year period.

27. specify, operationalize, and apply criteria for evaluating evaluation systems.

45. specify and operationalize criteria for evaluating evaluation systems.

37. apply a set of operationalized criteria to the evaluation of an evaluation system.

28. organize and administer a standardized testing program.

23. organize a standardized testing program.

9. administer a previously organized standardized testing program.

31. design and administer programs of research, development, instruction, and service for the improvement of evaluation.

35. administer programs of research, development, instruction, and service for the improvement of evaluation.

44. design programs of research, development, instruction, and service for the improvement of evaluation.

32. design, budget, arrange for, and support internal audits of evaluation systems.

13. design internal audits of evaluation systems.

61. budget, arrange for, and support internal audits of evaluation systems.

33. develop and disseminate clear and concise descriptions of evaluation systems that can be understood by the public.

33. develop clear and concise descriptions of evaluation systems that can be understood by the public.

18. disseminate descriptions of evaluation systems to the appropriate public through various means.

34. organize and administer a public information service in relation to evaluation information.

24. organize a public information service in relation to evaluation information.

I can . . .

53. administer a previously organized public information service in relation to evaluation information.

35. design and administer evaluation studies that monitor the implementation of projects.

6. design evaluation studies that monitor the implementation of projects.

4. conduct, given an appropriate design, evaluation studies that monitor the implementation of projects.

36. design and conduct a simulation study of alternative educational evaluation processes.

10. design a simulation to teach alternative educational evaluation processes.

51. conduct a simulation, given the materials, to teach alternative educational evaluation processes.

(8 Original Single-verb SAES Items)

8. describe the use of standardized tests for placement and diagnostic purposes.

12. provide stimulating leadership and direction to those who are serving in evaluation roles.

14. use audio-visual aids appropriately in making oral evaluation reports.

18. develop a schedule of reporting activities.

21. describe the dynamics of small group behavior.

25. design a sampling plan for field testing materials.

29. critique an evaluation report for its relevance to real decision problems.

30. interpret test results given in standard score format such as, grade equivalent, age equivalent, percentiles, deciles, or stanines.

ANSWER SHEET FOR SELF ASSESSMENT OF EVALUATION SKILLS (SAES)

DIRECTIONS: In each set of parentheses below, write the number which corresponds to your perceived degree of competence, interest, and importance for each of the items in the Self Assessment of Evaluation Skills (SAES). Write one number for each of the three scales for each item. Shortened definitions for each point on the scales are given below; for a more detailed description, please refer to the "Directions for Completing the Self Assessment of Evaluation Skills" which accompany these materials.

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest - area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest - area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevant to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP
1.	( )	( )	( )	11.	( )	( )	( )	20.	( )	( )	( )	30.	( )	( )	( )	44.	( )	( )	( )
4.	( )	( )	( )	13.	( )	( )	( )	21.	( )	( )	( )	31.	( )	( )	( )	46.	( )	( )	( )
5.	( )	( )	( )	15.	( )	( )	( )	22.	( )	( )	( )	37.	( )	( )	( )	47.	( )	( )	( )
7.	( )	( )	( )	16.	( )	( )	( )	23.	( )	( )	( )	38.	( )	( )	( )	48.	( )	( )	( )
8.	( )	( )	( )	17.	( )	( )	( )	24.	( )	( )	( )	39.	( )	( )	( )	49.	( )	( )	( )
9.	( )	( )	( )	18.	( )	( )	( )	26.	( )	( )	( )	40.	( )	( )	( )	52.	( )	( )	( )
10.	( )	( )	( )	19.	( )	( )	( )	27.	( )	( )	( )	42.	( )	( )	( )	55.	( )	( )	( )
	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP	56.	( )	( )	( )
																	COMP	INT	IMP

ANSWER SHEET FOR SEef ASSESSMENT OF EVALUATION SKILLS (SAES)

DIRECTIONS: In each set of parentheses below, write the number which corresponds to your perceived degree of competence, interest, and importance for each of the items in the Self Assessment of Evaluation Skills (SAES). Write one number for each of the three scales for each item. Shortened definitions for each point on the scales are given below; for a more detailed description, please refer to the "Directions for Completing the Self Assessment of Evaluation Skills" which accompany these materials.

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest - area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest - area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevent to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP
1.	( )	( )	( )	8.	( )	( )	( )	15.	( )	( )	( )	22.	( )	( )	( )	29.	( )	( )	( )
2.	( )	( )	( )	9.	( )	( )	( )	16.	( )	( )	( )	23.	( )	( )	( )	30.	( )	( )	( )
3.	( )	( )	( )	10.	( )	( )	( )	17.	( )	( )	( )	24.	( )	( )	( )	31.	( )	( )	( )
4.	( )	( )	( )	11.	( )	( )	( )	18.	( )	( )	( )	25.	( )	( )	( )	32.	( )	( )	( )
5.	( )	( )	( )	12.	( )	( )	( )	19.	( )	( )	( )	26.	( )	( )	( )	33.	( )	( )	( )
6.	( )	( )	( )	13.	( )	( )	( )	20.	( )	( )	( )	27.	( )	( )	( )	34.	( )	( )	( )
7.	( )	( )	( )	14.	( )	( )	( )	21.	( )	( )	( )	28.	( )	( )	( )	35.	( )	( )	( )
	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP	36.	( )	( )	( )
																	COMP	INT	IMP

ANSWER SHEET FOR SELF ASSESSMENT OF EVALUATION SKILLS (SAES)

DIRECTIONS: In each set of parenthese below, write the number which corresponds to your perceived degree of competence, interest, and importance for each of the items in the Self Assessment of Evaluation Skills (SAES). Write one number for each of the three scales for each item. Shortened definitions for each point on the scales are given below; for a more detailed description, please refer to the "Directions for Completing the Self Assessment of Evaluation Skills" which accompany these materials.

COMPETENCE

1. No Competence - completely unable to perform task
2. Minimum Competence - minimal performance with extensive study; able to hire knowledgeable consultant
3. Moderate Competence - moderate performance with minimal study; good performance with extensive study
4. High Competence - good performance with minimal study; exceptional performance with extensive study
5. Superior Competence - exceptional performance with little or no study

INTEREST

1. Negative Interest- area is repugnant to respondent
2. No Interest - no interest in area
3. Minimal Interest- area is somewhat interesting; other areas more desirable
4. Moderate Interest - area is genuinely interesting; other areas more desirable
5. High Interest - few areas more interesting to respondent

IMPORTANCE

1. No Importance - task almost never performed by respondent; considered irrelevant to professional performance
2. Minimal Importance - task performed occasionally; considered irrelevant to professional performance
3. Moderate Importance - task performed occasionally; considered relevant to professional performance
4. High Importance - task considered relevant, regardless of frequency of performance
5. Crucial Importance - task considered vital, regardless of frequency of performance

	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP
1.	( )	( )	( )	12.	( )	( )	( )	23.	( )	( )	( )	34.	( )	( )	( )	45.	( )	( )	( )	55.	( )	( )	( )
2.	( )	( )	( )	13.	( )	( )	( )	24.	( )	( )	( )	35.	( )	( )	( )	46.	( )	( )	( )	56.	( )	( )	( )
3.	( )	( )	( )	14.	( )	( )	( )	25.	( )	( )	( )	36.	( )	( )	( )	47.	( )	( )	( )	57.	( )	( )	( )
4.	( )	( )	( )	15.	( )	( )	( )	26.	( )	( )	( )	37.	( )	( )	( )	48.	( )	( )	( )	58.	( )	( )	( )
5.	( )	( )	( )	16.	( )	( )	( )	27.	( )	( )	( )	38.	( )	( )	( )	49.	( )	( )	( )	59	( )	( )	( )
6.	( )	( )	( )	17.	( )	( )	( )	28.	( )	( )	( )	39.	( )	( )	( )	50.	( )	( )	( )	60.	( )	( )	( )
7.	( )	( )	( )	18.	( )	( )	( )	29.	( )	( )	( )	40.	( )	( )	( )	51.	( )	( )	( )	61.	( )	( )	( )
8.	( )	( )	( )	19.	( )	( )	( )	30.	( )	( )	( )	41.	( )	( )	( )	52.	( )	( )	( )	62.	( )	( )	( )
9.	( )	( )	( )	20.	( )	( )	( )	31.	( )	( )	( )	42.	( )	( )	( )	53.	( )	( )	( )	63.	( )	( )	( )
10.	( )	( )	( )	21.	( )	( )	( )	32.	( )	( )	( )	43.	( )	( )	( )	54.	( )	( )	( )	64.	( )	( )	( )
11.	( )	( )	( )	22.	( )	( )	( )	33.	( )	( )	( )	44.	( )	( )	( )		COMP	INT	IMP		COMP	INT	IMP
	COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP		COMP	INT	IMP								

Commission for the Study of the Evaluation Profession

Evaluation Center / College of Education / The Ohio State University / Columbus, Ohio 43210

December 7, 1972

Dear Colleague:

Recently you received a letter from us requesting your assistance in a pilot test of the Self Assessment of Evaluation Skills instrument. To date, we have not received your completed questionnaire. Your response is most important in order to have an adequate critique of the items in the instrument.

In case you might have misplaced the previous mailing, we are enclosing copies of the materials we sent to you earlier. You will find another return envelope enclosed for your convenience. Your cooperation in completing and returning the materials will be greatly appreciated.

Sincerely,

Mary Anne Bunda

Mary Anne Bunda

Daniel L. Stufflebeam

Daniel L. Stufflebeam
Study Co-directors

APPENDIX E

Eight Logically-derived Categories of SAES Items

Category	Item Numbers
1. Knowledge of innovation in evaluation	1,2,9,10,15,21,29,30,35,58,67,80,82, 90,91,93,95,102,103,104,109,111,120, 131,132,134,135,136,153,178,183,184, 187,188,193,198,202,204,208,211,215, 222,225,226,231,233
2. Public relations	8,20,39,47,53,81,115,234
3. Data processing	128,141,154,156,185,201,203,210,213, 217,224
4. Educational measurement	34,55,66,76,83,97,100,105,106,107, 119,142,144,152,155,157,158,159,161, 165,167,169,176,177,181,190,192,194, 196,197,220,229,230,232
5. Evaluation administration	3,4,6,7,11,13,14,16,17,19,22,26,27, 28,32,36,37,38,40,44,45,46,54,56,59, 60,61,63,68,73,74,75,78,84,85,89,92, 96,98,101,108,112,113,114,116,121, 130,149,199,200
6. Relating evaluation to relevant disciplines	124,162,180,182,186,191,195,206,207. 214,221,228
7. Communications	5,12,18,24,25,31,33,41,43,49,52,57, 62,64,70,71,72,79,86,88,99,118
8. Research design analysis	23,42,48,50,51,65,69,77,87,94,110, 117,122,123,125,126,127,129,133,137, 138,139,140,143,145,146,147,148,150, 151,160,163,164,166,168,170,171,172, 173,174,175,179,189,205,209,212,216, 218,219,223,227

APPENDIX F

Item Means, Standard Deviations, & Inter-scale Correlations

ITEM OATA FOR SAES ITEMS
(Means, Standard Oeviations, Inter-scale Correlations)

In the table below, item data for the 234 items .n the SAES instrument (divided into the eight logically derived categories) are presented. Means and standard deviations for the competence scale are given since most other analyses described in this report pertain to competence only. In the inter-scale correlations, r_{ci} is the correlation between competence and interest, r_{cm} is the correlation between competence and importance, and r_{im} is the correlation between interest and importance.

1. KNOWLEOGE OF INNOVATION IN EVALUATION

	Competence		Inter-scale Correlations		
Item	Mean	Standard Oeviation	r_{ci}	r_{cm}	r_{im}
1	3.84	.76	.07	.13	.62
2	3.01	1.18	.44	.52	.68
9	3.58	.85	.39	.44	.70
10	3.70	.92	.52	.64	.70
15	3.16	1.08	.46	.54	.64
21	3.62	1.01	.49	.38	.63
29	3.55	.93	.47	.40	.68
30	3.09	1.13	.59	.46	.51
35	3.36	1.06	.46	.47	.55
58	3.51	.87	.49	.43	.62
67	2.96	.97	.44	.43	.69
80	3.35	1.01	.49	.57	.57
82	2.96	1.07	.49	.38	.66
90	3.36	.90	.6:	.41	.63
91	3.10	.95	.59	.45	.68
93	3.64	.92	.39	.30	.76
95	3.44	.91	.54	.40	.63
102	3.31	1.00	.60	.44	.63
103	2.87	1.01	.63	.54	.62
104	3.51	.91	.53	.46	.67
109	3.16	.92	.65	.30	.60
111	3.69	1.05	.58	.38	.60
120	3.03	1.06	.63	.63	.74
131	3.42	1.10	.70	.50	.73

I. KNOWLEDGE OF INNOVATION IN EVALUATION (cont.)

Item	Competence		Inter-scale Correlations		
	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
132	2.62	.96	.65	.57	.60
134	3.57	.88	.61	.61	.66
135	3.51	.87	.53	.39	.61
136	3.87	.82	.62	.54	.74
153	3.77	.94	.48	.36	.55
178	3.12	1.13	.64	.66	.75
183	3.68	.85	.59	.51	.73
184	2.79	.95	.50	.45	.68
187	3.26	.91	.73	.45	.56
188	2.89	.89	.59	.47	.64
193	2.73	.85	.66	.53	.74
198	3.51	.98	.64	.49	.59
202	2.71	1.00	.60	.70	.69
204	3.60	.78	.47	.40	.54
208	2.99	.93	.59	.40	.61
211	3.22	.95	.66	.50	.63
215	2.76	1.19	.60	.51	.74
222	2.18	.96	.55	.36	.51
225	3.10	.94	.63	.44	.58
226	2.81	1.03	.57	.43	.52
231	2.32	1.08	.60	.46	.72
233	3.09	.92	.64	.40	.64

II. PUBLIC RELATIONS

Item	Competence		Inter-scale Correlations		
	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
8	3.38	1.05	.62	.21	.45
20	3.96	.79	.54	.23	.32
39	3.49	.93	.58	.39	.45
47	3.74	.94	.33	.30	.54
53	3.55	.80	.60	.47	.50
81	2.77	.96	.34	.22	.54
115	3.09	1.09	.64	.46	.52
234	3.39	.95	.56	.38	.64

III. DATA PROCESSING

Item	Competence Mean	Standard Deviation	Inter-scale Correlations r_{ci}	r_{cm}	r_{im}
128	2.12	.87	.25	.16	.37
141	2.27	1.02	.36	.34	.31
154	3.17	1.09	.60	.50	.55
156	2.95	1.16	.65	.43	.38
185	3.03	1.36	.70	.31	.45
201	2.32	1.03	.56	.34	.40
203	2.10	.88	.48	.35	.49
210	2.83	.97	.66	.50	.60
213	2.43	.99	.57	.36	.52
217	2.60	1.02	.58	.27	.50
224	2.70	1.05	.54	.27	.56

IV. EDUCATIONAL MEASUREMENT

Item	Competence Mean	Standard Deviation	Inter-scale Correlations r_{ci}	r_{cm}	r_{im}
34	2.73	1.06	.35	.34	.55
55	3.49	1.00	.60	.46	.65
66	3.83	.98	.67	.46	.58
76	3.84	.89	.38	.35	.31
83	4.03	1.06	.55	.51	.75
97	3.64	1.13	.49	.43	.71
100	2.99	1.01	.52	.42	.74
105	4.18	.91	.35	.43	.63
106	4.09	.93	.42	.36	.55
107	3.84	.78	.40	.35	.63
119	4.00	.82	.37	.30	.65
142	3.45	1.09	.67	.51	.54
144	3.84	1.06	.52	.43	.40
152	3.95	.86	.46	.31	.63
155	3.66	.91	.68	.40	.54
157	2.79	.89	.62	.46	.60
158	4.26	.86	.53	.28	.46
159	3.22	1.31	.72	.70	.75
161	3.42	1.02	.61	.56	.70
165	3.88	1.14	.53	.36	.57
167	3.39	.88	.56	.55	.66
169	2.61	1.16	.76	.51	.77

IV. EDUCATIONAL MEASUREMENT (cont.)

	Competence		Inter-scale Correlations		
Item	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
176	4.05	.94	.57	.38	.52
177	4.39	.83	.44	.36	.42
181	3.29	.98	.64	.54	.59
190	3.45	.98	.49	.47	.63
192	3.27	.94	.45	.37	.65
194	4.29	.81	.42	.25	.43
196	3.79	1.08	.52	.32	.50
197	3.61	.93	.50	.52	.59
220	3.56	1.02	.60	.51	.75
229	3.87	1.00	.47	.45	.66
230	3.75	1.03	.62	.57	.70
232	3.43	1.03	.54	.50	.62

V. EVALUATION ADMINISTRATION

	Competence		Inter-scale Correlations		
Item	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
3	3.97	.87	.38	.32	.51
4	2.75	1.11	.49	.58	.64
6	3.05	1.01	.33	.48	.49
7	3.22	.99	.63	.48	.64
11	2.88	1.07	.34	.28	.46
13	3.06	1.24	.55	.60	.60
14	3.66	.96	.43	.41	.71
16	3.54	1.14	.70	.59	.76
17	3.08	1.17	.67	.49	.66
19	2.65	1.07	.56	.42	.42
22	3.43	.94	.51	.50	.59
26	3.49	1.05	.51	.35	.47
27	3.89	.93	.63	.38	.39
28	3.69	.83	.46	.41	.50
32	3.38	.93	.45	.35	.60
36	4.30	.78	.53	.45	.55
37	2.76	.95	.65	.59	.71
38	3.65	1.02	.52	.60	.54
40	3.58	1.07	.69	.38	.51
44	3.53	.97	.57	.57	.51
45	3.26	1.09	.43	.49	.47
46	3.49	.88	.52	.53	.53
54	2.77	.97	.65	.47	.65

V. EVALUATION ADMINISTRATION (cont.)

	Competence		Inter-scale Correlations		
Item	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
56	3.43	1.00	.66	.61	.67
59	3.17	1.14	.69	.60	.66
60	2.74	1.17	.70	.50	.65
61	2.89	1.13	.67	.38	.61
63	3.08	1.07	.50	.36	.68
68	3.44	.94	.63	.48	.54
73	3.10	1.12	.57	.55	.63
74	2.95	1.06	.59	.54	.54
75	3.66	.99	.54	.33	.55
78	2.79	1.12	.61	.48	.61
84	3.51	1.01	.44	.45	.70
85	3.75	.80	.57	.47	.71
89	3.59	.91	.50	.57	.63
92	3.45	.84	.61	.61	.72
96	3.11	1.06	.49	.44	.55
98	3.47	.87	.55	.48	.64
101	3.57	1.02	.68	.56	.63
108	3.66	1.03	.47	.60	.71
112	3.90	.82	.60	.39	.62
113	3.66	.84	.54	.39	.56
114	3.42	.96	.58	.47	.60
116	3.83	.77	.43	.44	.55
121	3.07	1.11	.56	.45	.68
130	3.63	.81	.61	.44	.67
149	3.36	.92	.69	.61	.61
199	2.99	1.06	.64	.54	.61
200	3.27	1.03	.52	.52	.67

VI. RELATING EVALUATION TO RELEVANT DISCIPLINES

	Competence		Inter-scale Correlations		
Item	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
124	1.97	.86	.44	.41	.61
162	2.94	1.04	.60	.47	.59
180	3.60	.88	.54	.51	.59
182	3.00	.84	.59	.55	.71
186	2.86	1.04	.68	.57	.59
191	2.69	.86	.59	.48	.62

VI. RELATING EVALUATION TO RELEVANT DISCIPLINES (cont.)

	Competence		Inter-scale Correlations		
Item	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
195	2.42	.79	.51	.36	.65
206	3.29	1.07	.62	.59	.62
207	2.51	.88	.44	.38	.66
214	2.65	.94	.54	.47	.57
221	2.34	.95	.58	.51	.61
228	2.18	.83	.64	.55	.62

VII. COMMUNICATIONS

	Competence		Inter-scale Correlations		
Item.	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
5	3.31	1.08	.47	.39	.61
12	4.07	.79	.35	.29	.69
18	3.08	1.11	.60	.50	.62
24	3.60	.94	.51	.27	.49
25	3.42	1.10	.52	.44	.70
31	3.21	1.18.	.49	.45	.53
33	3.10	1.03	.49	.37	.66
41	3.36	.99	.68	.68	.63
43	3.22	.96	.33	.31	.42
49	2.65	.94	.56	.48	.58
52	3.17	1.06	.58	.44	.57
57	3.26	1.02	.52	.34	.62
62	2.65	1.16	.56	.53	.58
64	3.53	.95	.56	.59	.66
70	2.73	1.03	.57	.49	.70
71	3.25	.71	.50	.47	.67
72	3.00	.89	.56	.47	.67
79	3.44	.85	.43	.37	.63
86	3.34	.98	.64	.48	.67
88	3.87	1.03	.53	.31	.55
99	3.38	.97	.49	.43	.55
118	3.44	.84	.51	.40	.58

	Competence		Inter-scale Correlations		
Item	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
23	2.09	1.02	.63	.36	.38
42	2.66	.91	.56	.46	.54
48	3.71	.95	.49	.30	.55
50	3.39	.92	.44	.54	.47
51	3.44	.97	.45	.31	.56
65	3.00	1.05	.47	.47	.63
69	2.51	1.11	.63	.44	.61
77	4.25	.76	.40	.31	.60
87	2.45	.97	.41	.16	.45
94	3.94	.86	.43	.10	.59
110	3.38	1.06	.65	.46	.51
117	2.95	1.00	.52	.32	.60
122	3.52	1.26	.53	.40	.60
123	2.68	1.28	.57	.53	.59
125	3.55	1.06	.64	.47	.60
126	3.38	.97	.58	.51	.61
127	3.94	.83	.47	.24	.48
129	2.49	1.19	.65	.36	.63
133	2.77	1.05	.50	.31	.59
137	3.19	1.01	.76	.56	.70
138	2.79	.98	.52	.48	.66
139	3.29	1.13	.73	.56	.69
140	3.22	1.19	.66	.55	.54
143	2.65	1.12	.66	.54	.63
145	2.97	1.04	.61	.51	.66
146	2.35	1.13	.66	.49	.55
147	2.90	1.12	.69	.57	.62
148	3.44	.99	.69	.68	.78
150	3.44	.84	.55	.43	.62
151	2.81	1.01	.63	.46	.60
160	3.00	.84	.50	.43	.58
163	3.56	1.02	.41	.39	.40
164	3.31	1.05	.66	.49	.64
166	3.16	1.05	.61	.39	.49
168	2.78	.98	.54	.44	.59
170	2.86	.98	.65	.54	.60
171	2.97	1.33	.69	.56	.64
172	3.13	1.04	.59	.50	.70
173	3.81	1.06	.64	.35	.50
174	2.40	.94	.62	.60	.65
175	3.19	.84	.49	.30	.56

VIII. RESEARCH DESIGN AND ANALYSIS (cont.)

	Competence		Inter-scale Correlation		
Item	Mean	Standard Deviation	r_{ci}	r_{cm}	r_{im}
179	3.21	1.02	.57	.43	.57
189	3.57	.92	.29	.37	.65
205	3.56	1.07	.68	.48	.62
209	2.60	1.15	.72	.43	.58
212	2.51	1.08	.65	.40	.59
216	2.92	.96	.49	.48	.63
218	3.92	.85	.47	.38	.67
219	3.83	.77	.41	.27	.59
223	3.04	.91	.59	.49	.61
227	2.20	.99	.71	.55	.59

APPENDIX G

Revised SAES Items

--Revised SAES Items

In general, all double verb items noted in the Item Contamination Study should be divided into two separate single verb items, as shown in Appendix O.

In addition, it is recommended that the following items be revised or dropped from the SAES instrument which appears in Appendix A.

Item	Recommendation
11	Revise to read "I can critique an evaluation report to determine the cost-effectiveness of the program (or project, materials, etc.) being evaluated."
23	Drop from SAES
37	Revise to read "I can project funding requirements (including core internal support, external support, and a fee structure for external evaluation services) for the evaluation system."
41	Revise to read "I can provide information which can be used to effectively disseminate a curriculum package."
45	Revise to read "I can formulate a budget for a specific evaluation study."
50	Revise to read "I can conceptualize a set of performance indicators that would form the basis for a school district's evaluation information system."
51	Revise to read "I can conduct a comparative study among extant curricular materials and a new curriculum package."
58	Revise to read "I can present a case for the evaluation of competing instructional strategies and suggest alternate evaluation methodologies which might be used."
60	Revise to read "I can project cost and political implications from evaluation reports that assess competing program strategies."
67	Revise to read "I can describe and analyze several major current programs of research, development, and instruction in evaluation."
69	Revise to read "I can discuss the assumptions underlying equal appearing interval scaling techniques and Thurstonian scaling techniques."

Item	Recommendation
81	Revise to read "I can design a video-tape presentation of an evaluation report."
92	Drop from SAES
103	Revise to read "I can describe a specific evaluation system to a group of evaluation theoreticians, relating the system to extant theoretical formulations in evaluation."
111	Revise to read "I can describe the difference between formative and summative evaluation in terms of information needs."
124	Drop from SAES
126	Revise to read "I can understand and judge the utility of projected data analyses."
134	Revise to read "I can generate criteria for evaluating a developmental instructional program such as 'Individually Prescribed Instruction'."
141	Revise to read "I can describe the basic structure of the FORTRAN computer language."
143	Revise to read "I can compare and contrast the definition of true score (from classical measurement theory) with the definition of universe score (in generalizability theory)."
157	Revise to read "I can discuss current developments in the field of educational measurement."
159	Revise to read "I can develop a general design (or test blueprint) to guide me in constructing a test."
162	Revise to read "I can analyze the relevance of demographic data to educational evaluation."
167	Revise to read "I can analyze data from open-ended interviews."
169	Drop from SAES or remove "jargon" from item
170	Rewrite as four separate items
178	Drop from SAES
181	Drop from SAES
182	Revise to read "I can incorporate major theories in decision-making into a discussion of their utility in educational evaluation."

Item	Recommendation
185	Revise to read "I--can carry through computer-based analysis of variance, multiple regression studies, and factor analysis using canned programs."
186	Revise to read "I can describe and analyze the present state of development of the educational evaluation profession."
189	Revise to read "I can conduct a literature review in a substantive area, write a report from that review, and present the report to a curriculum development group."
195	Revise to read "I can relate the major principles of information theory to current models of educational evaluation."
202	Revise to read "I can discuss advocate team methodology and contrast it with traditional site visit methodology."
208	Revise to read "I can set up a framework within which program personnel can generate solution strategies for specified problems which are both testable and creative."
218	Revise to read "I can arrange for the conditions under which data are to be gathered (including proper orientation for teachers)."
221	Revise to read "I can describe several evaluation theories, models, and practices from fields other than education."
228	Drop from SAES or rewrite as two items

APPENDIX H

A Partial Validation of the New Conceptualization of Evaluation Competencies

Mary Anne Bunda

Western Michigan University

A PARTIAL VALIDATION OF THE NEW CONCEPTUALIZATION OF EVALUATION COMPETENCIES*

Mary Anne Bunda **
Western Michigan University

The definition of a profession in terms of the universe of behaviors which comprise the activity in that profession is not new, as Tom Hastings has pointed out (in another paper presented in the same AERA session). There is, however, a unique aspect to the Universe of Evaluation Competencies developed in the Ohio State Model Training Project. The uniqueness is that we believe all individuals in the field have some competence across the range of skills; differences in areas of expertise may be reflected in a profile of those skills. The test of this assumption was planned through the development and refinement of the Self Assessment of Evaluation Skills (SAES) instrument.

The initial draft of the instrument was written with the desire to delineate the universe of behaviors by which the evaluation profession is characterized at all levels of involvement and expertise. The initial draft was divided into ten major categories, each of which is necessary to build evaluation as a profession in the view of the authors of the instrument. These categories are: (1) Administrative Leadership in Evaluation, (2) Research in Evaluation, (3) Development of Evaluation Methodology, (4) Instruction in Evaluation, (5) Implementation in Evaluation, (6) Development of Evaluation Systems, (7) Diffusion of Evaluation Developments, (8) Use of Evaluation in Educational Leadership, (9) Use of Evaluation in Teaching, and (10) Use of

---

*A paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, February 28, 1973.

**The author expresses appreciation to all Model Training Project staff whose work is reported in this paper; in particular to Ms. Evelyn Brzezinski, who designed the pilot study and has had significant impact on the direction of the research.

Evaluation in Curriculum Development. Tne hope was that this instrument could be used to study the members of the evaluation profession and thus aid those involved in training evaluators.

Although the evaluation profession nas been investigated several times in conjunction witn other professional educational roles (e.g., researchers), it has never been the topic of an in-depth study itself. Tne SAES instrument then, was to be used in a national survey of the members of the evaluation profession. A national study of those engaged in evaluation activities would yield data which compre the competencies of persons in one role of evaluation activities (such as director of an evaluation unit) with those of persons in another role (such as technical assistant to an evaluation project). Tne instrument, therefore, was constructed to include items from various content areas. It was noped tnat no one content area was explicated better tnan another.

Tnree major areas required further study before the national survey could take place:

(1) Tne content of the instrument (i.e., the items) had to be considered valid

(2) Tne scoring system or profiling plan had to be sound.

(3) Tne response format had to be valid and reliable.

Tne initial procedure was content validation of the items in the instrument. Tnis procedure nad two complementary activities. The first activity involved a search of extant literature which dealt with skills possessed by evaluators. Tne literature was searched for confirmation of the skills included in the draft items and for additional skills which were documented as necessary for successful evaluative operation. Tne success of the instrument was assumed to be based upon its scope. Redundancy of

skills or slight differences in the presentation of skills were not considered important factors. Rather, primary importance was placed upon scope and completeness.

The second activity in content validation involved requesting leaders in the field of evaluation to review the SAES instrument. These individuals were sent the instrument, asked to critique each item, use the scale that had been developed to report their own skill, and add items concerned with skills which had been omitted from the initial list. In this way two sorts of data were collected. The scope of the items was validated and somewhat enlarged, while the preliminary scoring system and response format were tested.

The preliminary scoring system followed the ten roles which were used to develop the instrument and which were described earlier. After data were collected in terms of the initial ten roles. several shortcomings of the scoring system were discovered. First, overlap among the scales was not consistent. Some roles were more inter-dependent than others. Second, the response categories which were constructed were not an exhaustive set. Consequently, a decision was made to develop a new scoring system based upon the results of a factor analysis of the instrument, to be performed after the data from the national survey were collected. Several interim scoring systems were used and one of them will be presented later in this paper.

Other major changes occurred in the instrument in its early developmental stages. Most significant among these changes was the development of several response scales. Initially, data were collected only in terms of the competence a person perceived himself as having in each of the skills. The definition of these response categories is in Table 1. Initially. it was assumed that this scale was robust enough to allow for data collection

from incumbent evaluators, from students in an evaluation training program, and from learning experiences. If the computerized guidance system which Dan Stufflebeam described (in a paper presented in the same AERA session) were to be implemented, all three types of data would be necessary. However, it soom became apparent that the scale lacked specificity for individuals and was completely inappropriate for gathering data about learning experiences. Besides, information concerning the present competencies of incumbent evaluators might not necessarily give good data about the training needs of the field, and therefore should not be automatically inputted into a computerized guidance system. Thus, the incumbent evaluator form of the instrument was revised to include two scales in addition to competence. The scales used in the pilot test are presented in Table 2. The importance scale was introduced to help delineate areas of skill which were perceived by practicing evaluators as important to specific roles in the field. The importance scale is a combination of frequency of use and criticality. These two factors jointly were believed to determine the necessity of a certain skill in any job. However, because the scale was not unidimensional, many respondents had trouble using it. In future studies with the instrument, the importance scale will have to be rewritten. The interest scale was developed because it perhaps would be useful to know the patterns of interest held by members of various roles in evaluation. This form of the instrument was administered orally to students in an evaluation training program and to a small group of professional evaluators.

Only a few minor changes in the wording of items took place after this test. The instrument was considered ready to be piloted on a group of incumbent evaluators from our mailing list. It was hoped that the mailing

list. It was hoped that the mailing list would comprise all individuals involved in evaluation activities throughout the country. Finding a list of such individuals proved to be no small task. An initial pool of names to be used in the full survey was the mailing list of members of Division H of AERA. Before the list arrived, stratified sampling was planned, the stratifying variable being jobs as indicated by mailing addresses. The list, however, turned out to have home addressed for 47% of the division members. This caused us to abandon any hope of stratification by job. The list also failed to include many people who were known evaluators. Two plausible reasons come to mind immediately. First, Division H is relatively new in the organization and members of long-standing in other divisions may have failed to affiliate with Division H. Second, since the title of the division is School Evaluation and Program Development, individuals in evaluation roles in labs, centers, or universities may have felt that Division H does not represent their interest. In order to have a mailing list which adequately represents all evaluation roles, therefore, additions were made to the division list. Several sources were used to enlarge the list: (1) the present mailing list for the Interfacter, a monthly newsletter published by the Ohio State University Model Training Project, which included evaluators from the Model Training Project consortium agencies, (2) personal address file of several evaluation specialists, and (3) two lists of evaluation consultants compiled by Bob Stake. The list that we used when drawing the sample for the pilot test, then, was composed of approximately 2,500 names and addresses.

Since the hope was to perform a factor analysis on the instrument after the full survey, a decision was made not to item sample in the pilot test

but rather to ask each individual to respond to each of the 234 items on each of the three scales. We realized that the length of the task would depress our response rate. Something had to be done to encourage respondents to fill out the instrument. First, we formed a "Commission for the Study of the Evaluation Profession" to endorse our study. The Commission was balanced with leaders in the field of evaluation by geographical area and by their present role. That is, we invited Commission membership from the public school sector and the university sector of the evaluation profession. Each Commission member was asked for a commitment of at least two days of his time; the two days of work devoted by the Commission members involved a critique (prior to the pilot test) of the survey plan and instruments to be used in the survey and a critique of the reporting format for the full survey. Permission was requested to list each Commission member's name on the cover letter stationery used in the pilot test.

The second strategy for improving our response rate involved a promise to respondents. In the full survey we planned to return to each respondent his score profile along with a booklet of mean profiles for each evaluative role. Since the scoring system was to be developed through a factor analysis of the full survey data, profiles would not be ready after the pilot test. Also, the data collected in the pilot might not be representative enough (because of a low response rate) to create norm profiles. Rather than use an a priori scoring system and send out score profiles without the appropriate norms to help interpret individual profiles, another sort of "pay-off" was promised to respondents. The Ohio State University Model Training Project at that time was publishing a monthly newsletter called the Interfacer which included spotlights on various consortium agencies, units within the

Evaluation Center, and students in the Model Training Project. Respondents who completed the pilot test forms were offered the opportunity to be placed on the mailing list to receive that newsletter.

In addition to the SAES instrument, a role description questionnaire was also included in the pilot test packet. The role description questionnaire was developed for use in the pilot test and, in a revised format, in the full survey. This questionnaire asked for various sorts of demographic and role data to help ascertain in what role group an individual belonged and to describe the norm groups in terms of academic experience, average length of time in position, etc. In addition, twenty-four educational roles were described and respondents were asked to place themselves within any one of the roles or to describe their present position. Hopes were to refine this questionnaire to a checklist for the full national survey.

The decision not to item sample led to another problem besides the possibility of a low response rate. The mailing cost for each instrument packet was fifty-six cents. Efforts were made to study the conditions under which we would get maximum response rate for the smallest mailing costs. Two major areas of mailing costs (reply format and follow-up format) were studied in the pilot test. A pilot test sample of 252 persons was randomly drawn from our population -- 120 persons with home addresses and 132 persons with business addresses. These subjects were randomly assigned to a six-cell matrix which represented three different reply formats (stamped return envelope, business reply envelope, and return envelope with no stamp) and two follow-up formats (postcard follow-up or follow-up letter plus questionnaire).

The results of the pilot test included the following:

(1) Of the 252 instrument packets mailed, 77 were usable responses, 8 were incomplete responses, 21 were blank or undeliverable, and 146 were nonrespondents. Thus only 42% of the questionnaires which were mailed ever found their way back to the Ohio State office. A telephone follow-up of a 15% random sample of the nonrespondents gave some leads to the causes of our low response rate. Of the sample nonrespondents, approximately 10% had changed positions and the forms had not been forwarded to them; approximately 20% said they were not evaluators and therefore did not complete the forms; approximately 20% indicated that they had completed or partially completed the instrument and returned it (however, we never received it). Forty percent of the sampled nonrespondents said they they did not complete the instrument because it was too long.

(2) The results of the tests of the manipulation of response and follow-up formats were:

(a) no significant difference in the return rate between persons who had received the questionnaire at home and at the office.

(b) no significant difference in rate of return between postcard follow-up and letter plus questionnaire follow-up either at the home or the office address.

(c) no significant difference in response rates of the return format of the home address respondents.

(d) a significant difference did occur in the response rates of the return formats at the office addresses. After one

follow-up, the business reply return rate was significantly higher than the no stamp return envelope at the .05 level. After the second follow-up notice the return rate elicited by business reply envelopes was significantly higher than the no stamp returns at the .01 level and the stamped envelope return was greater than the no stamp return at the .025 level.

On the basis of these statistics, decisions were made which would cut mailing costs during the full survey (e.g., business reply envelopes were chosen as most economical).

(3) The roles into which our respondents were classified are displayed in Table 3. As you can see, the majority of the respondents classified themselves into roles other than evaluation. The "other administrator" group includes deans, school superintendents, etc. The "other university professors" includes professors of, for example, educational administration, curriculum, and sociology. The group labeled "other" includes program and product developers, psychologists, secondary school teachers, and institutional researchers.

It was hoped that the results of the pilot test could also include some typical profiles of several of the role categories. However, because the roles were so varied, any one role had very few individuals in it. Therefore, profiles drawn on the basis of any one group would be misleading. However, a sample profile of skills is presented in Table 4 for illustrative purposes. The two groups represented are administrators of evaluation or research and researchers or statisticians. The profile scoring system was

developed by sorting the SAES skills into a priori categories. This sample profile is presented to illustrate what the results of the national survey might look like. Examples of the type of items in each of the categories are listed in Table 5. The sample profiles are drawn for the competence scale only. It is meant to be illustrative of the differentiation of roles on an evaluation team in terms of competence

Further analyses of the pilot test data are being completed at The Ohio State University. Included in these analyses is a study of item contamination. As you probably noticed from the sample items, some of the SAES items contain two verbs, such as "to design and implement". To determine if responses to double verb items differed from responses to items which contain a single verb, a small sub-study was conducted. Each of the 28 items in SAES which contain a double verb was divided into two single verb items, and three experimental instruments were created: 28 double verb items, 56 single verb items, and 28 single verb items which are a random half of the total 56 single verb items. The complete analysis of these data has not been completed, but will appear in the final report of the Ohio State Model Training Project.

ORIGINAL SCALE

1. No Competency

   Respondent has no training base from which to work and would be completely unable to perform this task.

2. Minimum ComPetency

   Respondent has had some training and experience in this area, but does not feel confident enough to perform the task without further training or complete supervision by someone knowledgeable in this area.

3. Moderate Competency

   Respondent has had fairly thorough training and experience in this area and could perform the task moderately well, but his skills are not up to date and many other people in the field have much more expertise in this area.

4. High Competency

   Respondent has a good background in this area and is completely comfortable in performing this task. Some people in the field however, still possess more expertise in this area.

5. Superior Competency

   Respondent is completely confident in performing this task and has done creative work in this area. (S)He is one of the most skillful persons in the profession in this area.

SCALES USED IN PILOT TEST

Competence Scale

1. No Competence - Respondent has no base from which to respond and would be completely unable to perform this task, or does not understand the terminology.

2. Minimum Competence - Respondent could perform this task minimally well only with extensive study. He does, however, know enough about the area to hire a knowledgeable consultant.

3. Moderate Competence - Respondent could perfom this task moderately well with minimal study and well with extensive study.

4. High Competence - Respondent could perform this task well with minimal study and exceptionally well with extensive study.

5. Superior Competence - Respondent could perform this task exceptionally well without study.

Interest Scale

1. Negative Interest - Respondent finds this area repugnant.

2. No Interest - Respondent has absolutely no interest in this area.

3. Minimal Interest - Respondent finds the area somewhat interesting but can generally find more enjoyable activities to engage in.

4. High Interest - Respondent has genuine interest in this area but there are professional activities he finds more enjoyable.

5. Superior Interest - There are few professional activities that the respondent would rather engage in.

Importance Scale

1. No Importance - Respondent almost never performs this task and considers it irrelevant to his professional performance.

2. Minimal Importance - Respondent occasionally performs this task but considers it irrelevant to his professional performance.

3. Moderate Importance - Respondent occasionally performs this task and considers it relevant to his professional performance.

4. High Importance - Respondent often performs this task and considers it relevant to his professional performance.

5. Crucial Importance - Ability to perform this task regardless of its frequency is vital to the respondent's successful functioning.

Table 4

I Knowledge of Innovation in Evaluation
M = 46

I Public Relations
M = 7

I Data Processing
M = 11

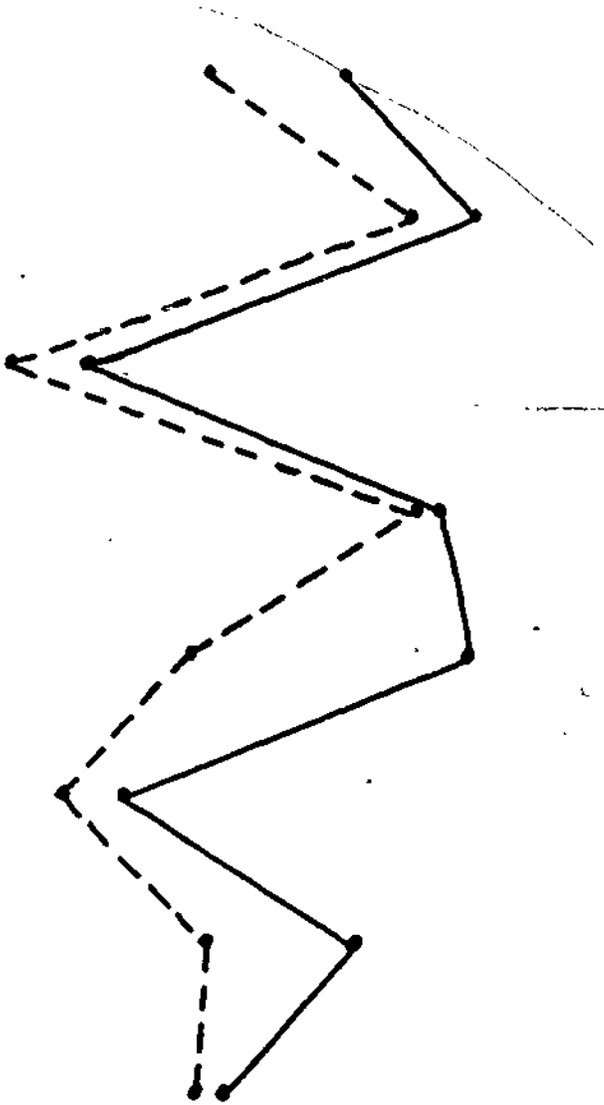
V Educational Measurement
M = 34

V Evaluation Administration Skills
M = 50

II Relating Evaluation to Relevant Disciplines
M = 12

II Communications
M = 22

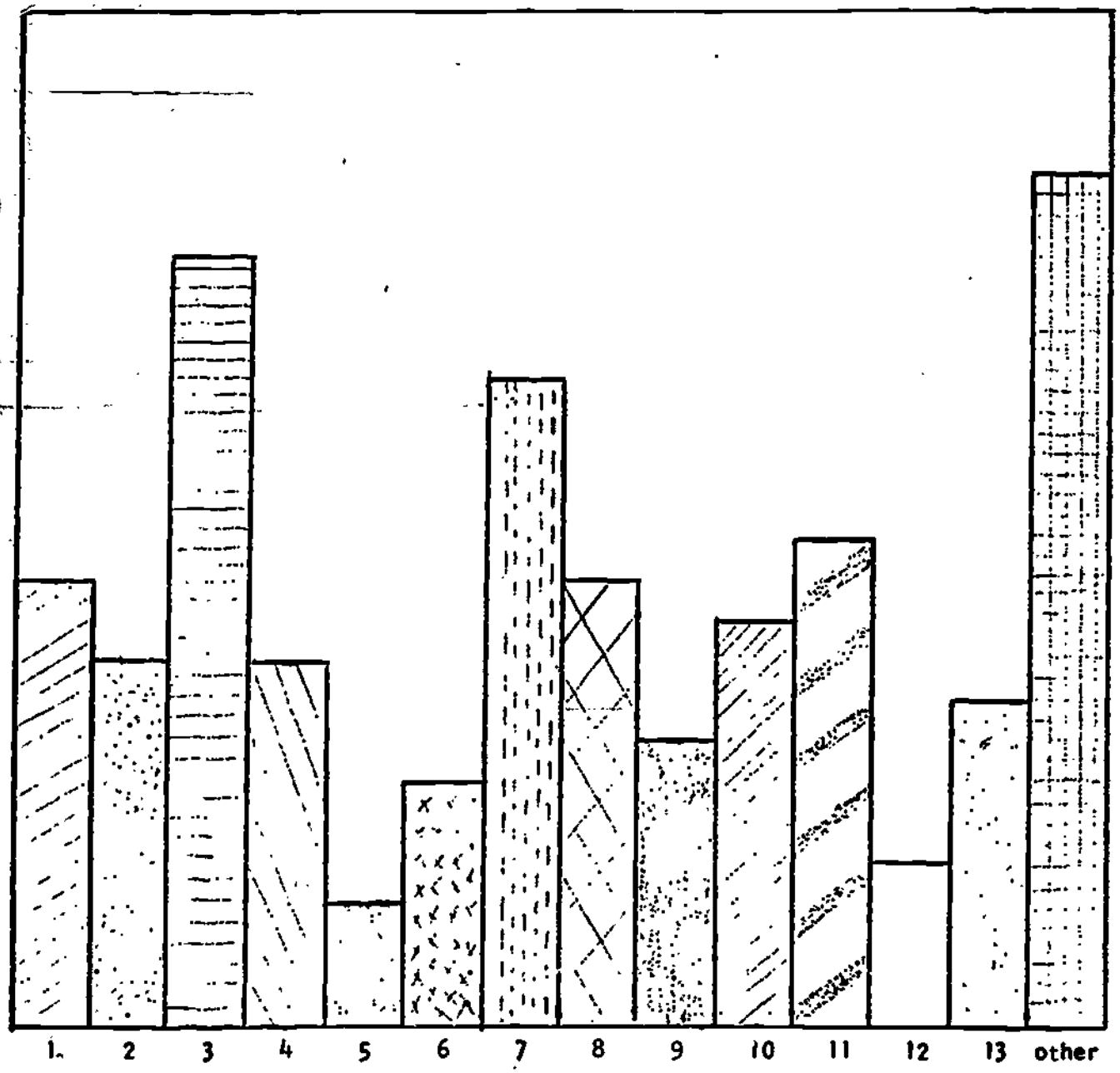
II Research Design Analysis
M = 51



N = 15 – – – –Researcher
N = 15 ————Administrator
M = number of items in category

1 2 3 4 5

ROLE DESCRIPTION OF RESPONDENTS



1 administrator: research

2 administrator: evaluation

3 other administrator

4 professor: research courses (e.g., design, statistics)

5 professor: evaluation courses (e.g., theory, design)

6 professor: development courses (e.g., media utilization, instructional design)

7 other college professor

8 specialist: evaluation design

9 specialist: evaluation implementation

10 consultant: evaluation

11 researcher

12 statistician/data analyst

13 evaluator of Federal projects

117

Table 5

ITEM EXAMPLES

I. Knowledge of Innovation in Evaluation

1. I can compare and contrast instructional research and evaluation.

58. I can present a case for the evaluation of competing instructional strategies and suggest alternate methodologies which might be used.

93. I can design mechanisms to collect judgmental information concerning the worth of a curriculum package.

211. I can select, organize, and lead groups of professionals to generate alternative program strategies and project designs.

II. Public Relations

47. I can use audio-visual aids appropriately in making oral evaluation reports.

81. I can design an evaluation report to be presented by video-tape.

234. I can describe and analyze user attitudes toward evaluation.

III. Data Processing

154. I can design and develop a data bank.

156. I can develop forms and procedures for managing the data processing operations of an evaluation system.

224. I can design an information system according to which data are to be coded, stored, and retrieved.

IV. Educational Measurement

34. I can list what audio-visual materials are available for the training of evaluators.

119. I can write forced-choice and free-choice test items.

176. I can select a test that has been normed on an appropriate group.

V. Evaluation Administration

4. I can marshal political support for evaluation activities.

96. I can develop a PERT network.

200. I can delineate evaluation authority and responsibility within the agency's organizational structure.

VI. Relating Evaluation to Relevant Disciplines

124. I can write an analysis of the relevance of economic theory to educational evaluation.

207. I can relate the major research strategies and theories of sociology, political science and economics to evaluation problems.

228. I can state the basic principles of value theory and utility theory and compare and contrast the relevance of these fields for theory development in educational evaluation.

VII. Communications

5. I can prepare a 20-minute slide-tape presentation on problems in evaluation.

70. I can organize and administer an editorial service in relation to evaluation reports.

118. I can develop and disseminate clear and concise descriptions of evaluation systems that can be understood by the public.

VIII. Research Design Analysis

94. I can design controlled research studies of small units in a curriculum.

189. I can conduct a literature review, write it up, and present it to a curriculum development group, irrespective of substantive area.

227. I can defend the choice of an oblique rotation or an orthogonal rotation in a factor analysis study.