DOCUMENT RESUME

ED 091 43C TM 003 652

AUTHOR	Hood, Paul D.
TITLE	Current Activities in the Design and Development of Development, Dissemination and Evaluation (D,D&E) Competence Catalogue and Assessment Instruments.
PUB DATE	[Apr 74]
NOTE	27p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	Certification; Decision Making Skills; Educational Development; Educational Objectives; *Educational Researchers; Evaluation; *Evaluation Techniques; *Professional Training; *Research Skills; Taxonomy
IDENTIFIERS	*Far West DD and E Assessment System; Performance Based Programs

ABSTRACT

A description of the current activities of the Far West Consortium for Development, Dissemination and Evaluation (DD&E) in developing a competency-based assessment system for educational research and development (R&D) professionals is presented. The taxonomy of competencies and the battery of assessment instruments being developed will initiate competency based education for R&D professionals. A problem faced by the consortium is the need for a concrete definition of what competencies are needed for specific levels of professionalization. One of the major design concepts of the assessment system is that it be decision oriented; that is, that the system be concerned with the utility and feasibility of competence based decision making. Four basic assessment methods have been selected: ratings, job knowledge tests, job sample tests, and product ratings. The reliability and validity of the assessment instruments are still being computed. In deriving the taxonomy of educational DD&E competence statements, the project worked with a small sample of subjects who may not be representative. The author hopes this work may serve as a stimulant to more research in this area of R&D competency assessment. (MLP)

CURRENT ACTIVITIES IN THE DESIGN AND DEVELOPMENT OF DEVELOPMENT, DISSEMINATION AND EVALUATION (D,D&E) COMPETENCE CATALOGUE AND ASSESSMENT INSTRUMENTS

Paul D. Hood

Far West Laboratory for Educational Research and Development

Paper Presented at the Annual Meeting of the
American Educational Research Association
Chicago, 1974

CURRENT ACTIVITIES IN THE DESIGN AND DEVELOPMENT OF DEVELOPMENT, DISSEMINATION AND EVALUATION (D,D&E) COMPETENCE CATALOGUE AND ASSESSMENT INSTRUMENTS

by Paul D. Hood

INTRODUCTION

My presumption is that most of the persons attending this session come with either a general interest in competency-based education or a specific interest in the assessment of R&D personnel. Like any competence-based educational program, the Far West Consortium for DD&E Training had to face the problem of how first to derive and then to assess the competencies on which the program was based. This is an immensely more difficult and challenging problem than we first suspected. To date, our accomplishments are quite modest. For those whose interests are in competence-based teacher education, I believe our methodological approach may be of interest.

For those of you with a more direct interest in assessment or even certification of educational R&D professionals, let me preview my conclusions. I believe the competence catalogue and the battery of assessment instruments we are developing will give us a start toward competence-based education for R&D professionals. Indeed, I believe our instruments and methods will be quite useful for any decision maker (student, instructor, employer) who is concerned with "personnel development," that is with counseling, guidance, training, job assignment, and the like. However, I have strong reservations about how far we are currently able to go toward professional "certification." These reservations are based on three immense deficiencies. First, we don't have a good "front-end" analysis, to borrow Susan Markle's term, with reasonably precise competence statements and a practical concensus of what competencies the educational DD&E professional needs. Second, we are far from having adequate instru-

mentation that will allow us to assess these competencies. Finally, we have virtually nothing in the way of credible validation data. The Far West project is addressing all three of these problem areas, but I'm certainly not going to claim remarkable accomplishment--just some progress.

THE FAR WEST DD&E ASSESSMENT SYSTEM

Although competence-based professional education was not so apparent in 1970 when we first designed the DD&E Training Program, it has emerged as a major movement. One of the most perplexing problems faced by this movement is the definition and assessment of competencies. Stanley Elam (1971, p. 21) calls it the overriding problem before which other problems pale to insignificance. David Krathwohl (in Merwin, 1973, p. v) states bluntly that one can predict that performance-based teacher education is certain to fail in reaching its ultimate objectives if it continues on its present course. And this failure will be caused by the almost complete lack of attention given to the assessment of competencies..."Until and unless some real progress is made on resolving the problems of instrumentation and measurement, PBTE will go down in the history books as one more bandwagon in the long line of over-simplistic solutions for complex problems."

I believe that Elam and Krathwohl are right. Despite years of psychometric R&D, the technical assessment tasks encountered by any competence-based program are still formidable. Moreover, the logical and philosophical issues are demanding. Currently, there seems to be no commonly-accepted approach to developing competence assessment instruments; in fact there is not even any commonly-accepted definition of "competence." Schalock and Thomas (1973) have made a useful distinction between two meanings commonly employed. One equates competence with the

mastery of knowledges and skills assumed necessary to perform a particular function. The second holds competence to be the demonstrated ability to bring about outcomes specified in a given job description.

Criterion Levels

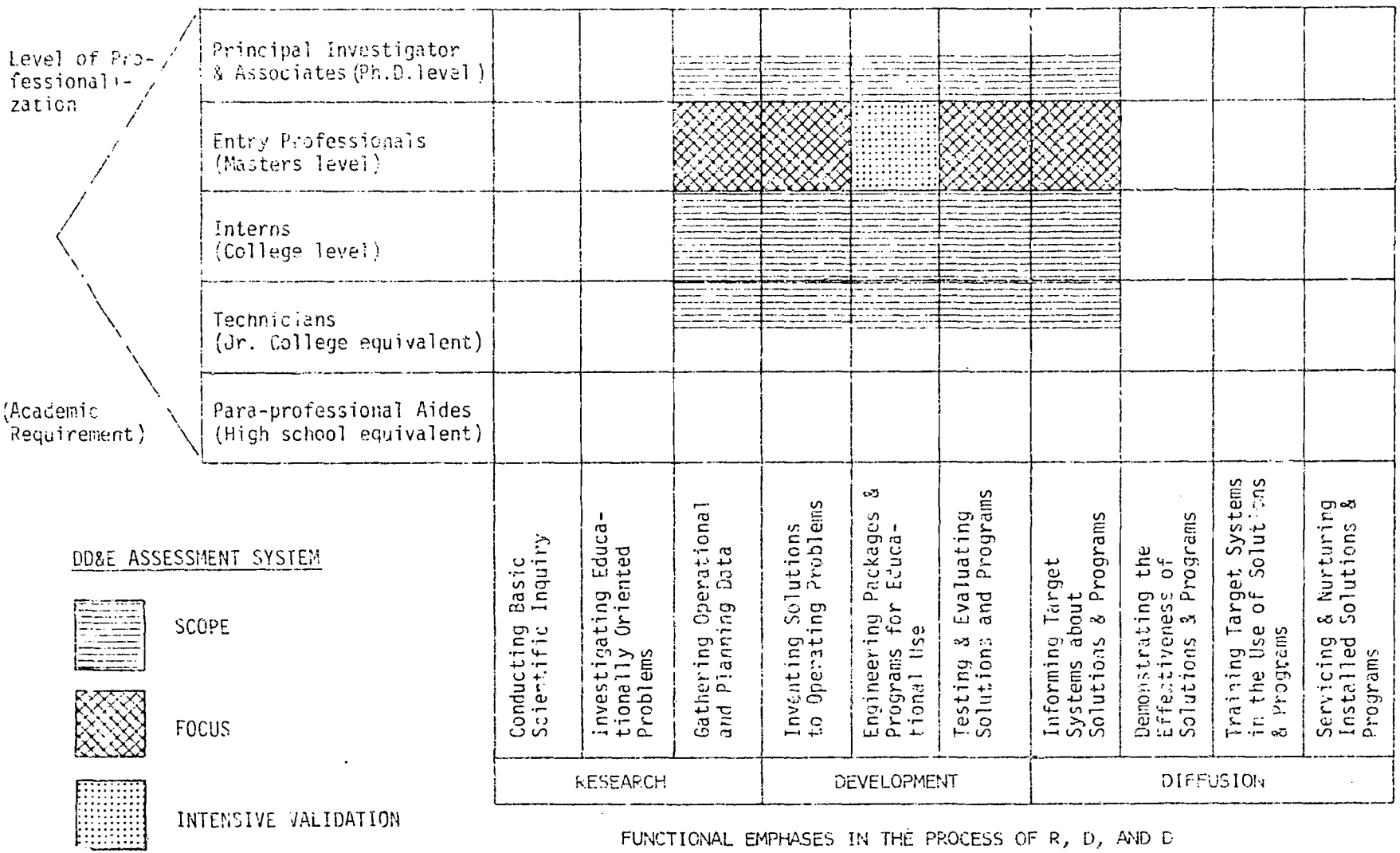
Richard Turner (1971) has provided a finer discrimination with six criterion levels. The six levels range from demonstration of mastery of knowledge and understanding (in level 6), demonstration of skill attainment in simple training and laboratory conditions (in levels 5 and 4), behaviors in actual conditions (in level 3), to evidence of short- and long-term pupil change (in levels 2 and 1). As we ascend Turner's criterion ladder, the problems and costs of assessment mount rapidly.

In our DD&E assessment system, we intend to provide tests at Turner's three lowest levels, namely tests of knowledges, and performance of skills in job sample tests and simulations. Assessment at level 3, behavior in actual conditions, is provided only through ratings by supervisors. Thanks to recent work by Popham and others (1974), we may someday have instrumentation appropriate for short- and long-term effects of a developer's products on target audiences in levels 1 and 2. However, since test development and data gathering costs appear so high, we are not even proposing assessment at these criterion levels in the present project.

It may be helpful to look at the competence content area addressed by the Far West project. (See Figure 1, adapted from Clark and Hopkins, 1969, p. 14.) Figure 1 locates the scope, focus, and area of concentration. Our focus is on competencies required of entry-professionals (masters degree level personnel) in the area of development, with some spill-over into immediately adjacent areas

Figure 1

ASSESSMENT INSTRUMENT DEVELOPMENT PRIORITIES FOR THE FUNCTIONAL COMPETENCE DD&E TRAINING PROGRAM



FUNCTIONAL EMPHASES IN THE PROCESS OF R, D, AND D

of research and diffusion. The reason for this focus is that it corresponds to the area addressed by the training materials we are developing. Figure 1 indicates that the scope--the area where our assessment materials may be useful--extends to a wider range of professionalization levels. Simply stated, although the content is the same, our assessment instruments may provide useful information about levels of competence which are below or slightly above those considered appropriate for DD&E entry-level professionals. Finally, Figure 1 indicates that we intend to make a concerted effort at rigorous instrument development and validation in the more restricted area of "engineering packages and programs for educational use."

Figure 2 indicates that the project, in its attempt to derive curriculum and instructional objectives, has drawn on three data bases, the AERA Task Force on Training (Worthen, et al., 1971, 1973), the Oregon Studies in Educational RDD&E (Schalock et al., 1973) and the Far West Consortium's task analyses (Hood, et al., 1970). Parallel development and field testing of instructional modules and of

tems provides field test data on students as well as expert review of con-

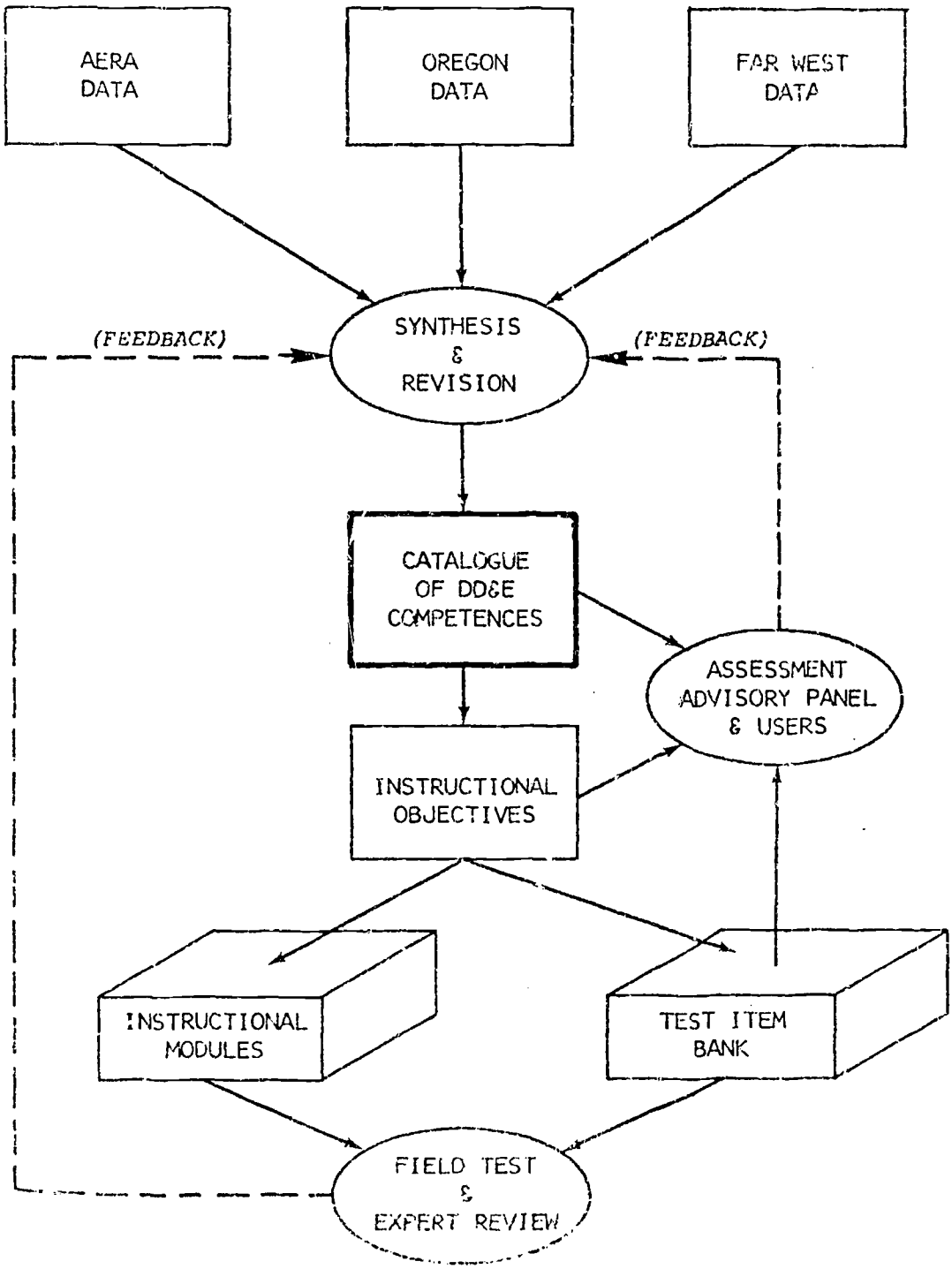
is experience is fedback in revision cycles resulting in modification of both the instructional resources and the assessment instrumentation.

Focusing now on the assessment side of the project, we see that four subtasks have been defined:

1. the creation of a DD&E Competence Catalogue,
2. the development of a diagnostic test battery,
3. the development of a competence assessment battery, and
4. the documentation of the project's derivation, development and validation methodology.

Figure 2

Derivation and Validation Linkages
for a Catalogue of DD&E Competencies
& Competence Assessment Instruments.



Design Concepts

One of the major design concepts in our approach is that the assessment system be decision-oriented. We are not so much concerned about tests and their psychometric properties as we are about the utility and feasibility of competence-based decision making. Who are the decision makers in a competence-based educational system?

First, there are the students. They are concerned with questions regarding status and progress. They want to know what competencies they have already mastered; what their level of proficiency is vis a vis training exit or employment requirements; and whether they should repeat a unit of instruction or proceed to the next unit.

Second, there are the instructors. They want to know the students' current levels of accomplishment and skills; what further training should be planned; what progress is being made; and whether in fact exit mastery is attained.

A third group are the employers. They are particularly interested in the match between a prospective employee's attainments and the job requirements. Does the person qualify for a particular job or will further on-the-job training be needed? Given valid and detailed information, employers may restructure the work for more effective use of their employees' current levels of competence or possibly may encourage more on-the-job training or inservice education.

Finally, there are those of us who are concerned with the development and evaluation of competence-based training programs and resources. We need to know about the range of entry-level knowledges, skills, and sensitivities. Do specific instructional materials and methods facilitate attainment of particular competencies? How can instructional materials and methods be improved? How can

credible evidence of program performance sufficiently impressive to persuade potential users to adopt the program or to hire its graduates be provided?

These are some of the questions that students, instructors, employers and developers may ask. How are the answers to be provided?

I have already indicated that one of our major design concepts is that the assessment system be decision oriented. This in turn implies that we be concerned with utility, usability, validity, and reliability; and in that order. If our instrumentation does not lead to better decision making, for all the decision makers, there is little chance that the assessment system will actually be implemented and even less chance that it will be maintained. So, above all, the system must be useful. It must have apparent and real utility in helping users make better decisions. The usability of a system means that time, costs, psychological threat, etc., must be minimized while maximizing the information obtained for each stakeholder--student, instructor, potential or actual employer. In our opinion, the requirements for utility and usability transcend those for validity and reliability. Obviously, there must be some non-trivial level of validity and reliability. But if one takes a decision-theoretic approach, it is the utility of the decision that is paramount. The major issues are the risks of error and the costs of those errors. The challenge for us has been how to configure a set of information gathering instruments and processes in a form that would lead to their acceptance and use and also provide for efficient organization of information that would have an actual bearing on decisions.

Three important and highly practical, technical points are that: 1) we are dealing with a classification model; 2) we are dealing with a multiple-sequential decision model; and 3) a multiple-sequential decision model may be best handled with a Bayesian statistical decision approach.

Many of our assessment approaches are based on selection models where we attempt on the basis of a one-time assessment to decide whether a student should be selected into or out of a training program, passed in a course or certified for graduation. Although selection remains an aspect of competence-based programs, the assessment problems really deal more with classification. The instructor, the student and the employer have different perspectives and the majority of their decisions are not simple "go-no-go" decisions. More often, it is a matter of deciding how well prepared, which job, what kind of a career, how much and what kind of supervision, what potential for advancement, what kind of instructional resources, etc.

The competence-based, individualized approach provides the need and the opportunity to make a sequence of tests and decisions. Few of them are totally irreversible. This is a fortunate situation since few of the measures which are feasible provide highly reliable or highly valid measures when used singly or on one occasion. We know from decades of personnel research work that interrater reliabilities of complex performances and products may not go much above .6 or .7 in actual practice. And validity correlations of a specific predictor to a specific criterion are often in the .2 to .4 range. While this situation is tolerable when dealing with groups of persons, it becomes less so when dealing with decisions specific to one person. However, if we take a relatively large number of measures <u>over time</u> and employ a <u>variety</u> of methods, the cumulative sequential decision process itself can attain a much higher reliability and validity with correspondingly lower classification errors and costs.

We have known for some time that, at least in theory, the Bayesian statistical approach was an attractive alternative to the classical approach, since it is rarely the case that any decision maker has a flat prior expectation with

total ignorance regarding the probabilities of alternative outcomes. Recent contributions by Brown (1969), Ferguson and Novick (1974) and others, suggest that we shall be able to provide relatively simple procedures for all real decision makers, whether they are students, instructors, employers, developers or even federal sponsors, to arrive at more effective decisions through fuller use of available information.

For the above reasons, we have had to focus on the idea of developing a highly flexible, inexpensive and practical assessment system rather than a simple collection of test instruments. Indeed, this perspective has caused us to be far more concerned about helping decision makers to use information more effectively than in simply developing more reliable and valid measurement devices. Please understand that I am not underrating the need for better measurement. I am simply asserting that from a very practical point of view, the crucial problem in competence-based education is to get people, including instructors and students, to want and know how to use assessment information. Until this happens, we may have technically impeccable but unusable competence assessment instrumentation. Hence, we have been concerned with making our procedures attractive, inexpensive, easy to use, easy to understand, face valid, useful and meaningful.

Assessment Methods

After review of a number of alternatives, we have selected four basic assessment methods: ratings, knowledge tests, job sample tests, and product ratings. Although these four methods will be augmented by biographical information, questionnaires, interviews, and other sources of information, these four methods constitute the foundation for assessment.

Ratings. We have found that while ratings by students, instructors, peers, and work supervisors are useful; possibly only the student rating may be feasible in the preservice program. Although subject to well-known types of error, extensive experience in the use of ratings indicates that if well-designed and properly employed they can yield modestly reliable and valid information at very low cost. After trying several formats we are currently working with a seven-point performance-based rating scale which permits relatively unambiguous determination of observable or easily inferable levels of performance.

Job knowledge tests. Initially, we attempted to develop highly objective and easy-to-score job knowledge tests. However, we have found it difficult to write items testing for comprehension, ability to make applications or evaluative judgments with completely objective formats. Our general approach now is to use essay and short-answer written questions, which call for more complex constructed responses on the part of the testee and also, unfortunately (?), for greater thought and attention on the part of the grader. This is one area where we have accepted increased grading costs in the hope of obtaining greater meaningfulness and validity.

Job sample tests. Ratings and job knowledge tests are relatively easy to develop and can cover a lot of "competence territory" in relatively short time. Job samples are something else. Generally, they are quite time consuming and expensive to administer and to score. Moreover, it is quite difficult to avoid a certain degree of situational specificity which may significantly limit the generality of results. Currently we are experimenting with a simulation test which consists of a series of separately scorable but logically related job samples.

Product ratings. Since many DD&E activities produce tangible products or by-products in the form of plans, outlines, scripts, or collections of bibliographic references, the systematic evaluation of these products provides an obvious and relatively inexpensive source of information. Allowance for the conditions in which the products are produced (e.g., outside assistance, time available) must of course be made if the products are not produced under controlled test conditions.

Actual Use

Our concerns for costs and practical use, have led to a sequential strategy in which the least expensive devices such as self-ratings are used most extensively, but with cross checking against more expensive sources of information such as knowledge tests and job samples. An item sampling strategy, coupled with a Bayesian decision approach in which test items are selected for their relative potential in reducing the decision makers uncertainty, is being developed.

The various types of instruments and their use are displayed in Figure 3. First note that we have created an artificial dichotomy in illustrating a difference between diagnostic use and assessment use. In the earlier stages of the program the emphasis is primarily on program planning, guidance, and counseling. As the student progresses, the emphasis tends to shift to concerns about completion of modules, attainment of objectives, credit for attainment, and ultimately to graduation and competence certification.

Comparable data, and sometimes identical or parallel test items and instruments may be employed for both uses. A student's file is "opened" by recording pertinent biographical, academic, and work experience data from the student's application form. This may be augmented by instructor interview data.

Figure 3

DD&E Assessment Instrument Use

INSTRUMENT TYPE	DIAGNOSTIC USE: °Guidance °Counseling °Program Planning	ASSESSMENT USE: °Progress Assessment °Credit-by-examination °Certification of Mastery
1. Ratings		
a. Self-Rating	Q1, Q2	Q3
b. Supervisor Rating	Q1, Q2 (if applicable)	Q3
c. Instructor Rating	---	Q3
2. Knowledge Tests	Module Pre-tests ---	--- Module Post Tests Development Series Knowledge Tests
3. Job Samples	(Available)	Instructional Module Job Samples Development Simulation
4. Product Ratings	(Available)	Development Products
5. Other Information	Biographical Data	(same)
	Interview Data	(updated)
	Academic & Work Data	plus program progress information

The first structured instrument the student encounters is a Q-sort (Q1) of 72 self-rating items which has been designed to reveal a competence profile on the three DD&E functional contexts (development, dissemination, and evaluation) and on six process skills (analyzing, planning, production, evaluating, collecting and organizing information, and communicating). On the basis of this information taken together (remember we are employing a sequential, Bayesian, item sampling strategy), the instructor may decide to probe areas where further information is needed--perhaps, because the students' self appraisals in those areas suggest discrepancies in either being higher or lower than expected. The least expensive alternatives available are: (1) interviews, (2) use of a second tailored Q-sort drawn from the rating item pool (Q2) or the use of one or more of the module pretests. In some instances, supervisors' ratings, job samples or product ratings may be available. However, their use in diagnostic situations would be unusual.

As students progress through the program they will encounter up to twenty-three end-of-module knowledge tests. And for those modules calling for the attainment of demonstrable skills, there may be job sample tests. The students may also produce a variety of rateable products as part of application projects. In some instances, they may have an internship where a valid job supervisor rating can be obtained.

At the end of the program this cumulative file of information can be augmented by a third self rating (Q-3). Because of the flexible item sampling format, this third self rating may be tailored to probe areas corresponding to the student's own program objectives. When relevant, supervisor or instructor ratings may also be obtained. Finally for the development series only, we shall also have a more comprehensive knowledge test and job sample (simulation) test.

Validation

Now a word about validation. Our approach is from three directions. First we are using a panel of experts, including DD&E work supervisors, to make judgments about face validity and relevance. Second, we are employing the multitrait, multimethod approach (Campbell and Fiske, 1959) to establish convergent and discriminant concurrent validity. This second approach requires that there be statistically significant correlations among several measures (e.g., ratings, knowledge tests, job samples) purporting to measure the same competence (trait) and that "off-diagonal" correlations involving the same methods, but different competencies be smaller than correlations among different methods on the same competencies. Finally, we are requiring that the measured differences pre- to post-training or between groups known to have markedly different competence levels be significant and of practically meaningful magnitudes.*

Regarding our progress, it is decidedly uneven. We have developed two mini-batteries involving ratings, knowledge tests, and job samples. Both batteries gave us encouraging reliabilities and convergent validities, but neither battery did well on discriminant validity except for the self and supervisor ratings. We have made good progress in developing the diagnostic Q-sort. Its acceptance by students and instructors is quite encouraging and the preliminary statistics are generally good. There is very high internal consistency for the several competence scales and evidence that the instrument detects meaningful differences in competence profiles. The sort-resort Q-correlation reliabilities are lower than we would like (.6 to .7), but the correlations with supervisors sorts are almost as good (.5 to .6).

---

* This last test is related to the convergent validity test, but is more stringent.

I wish I could be more positive about our knowledge tests. Some of our early attempts were embarassingly bad--mainly because of a preponderance of rote learning items. I think that we have better items now. Since we are still in data collection, I don't know what the results will look like. Currently, our big push is on the development of the simulation test for the Development Series.

A CATALOGUE OF COMPETENCIES IN EDUCATIONAL DD&E

One of our project objectives is to produce a catalogue of educational DD&E competencies. Methodologically, this effort addresses five related problems. They are:

1. How to create a classification scheme or taxonomy to organize competence statements.
2. How to articulate statements with or derive statements from evidence which supports their validity and relevance.
3. How to phrase competence statements efficiently.
4. How to articulate statements with available instructional and assessment resources.
5. How to design the catalogue so that it will be useful to various audiences; including developers of instructional programs and materials, developers of assessment instrumentation, researchers concerned with R&D performances and R&D functions, trainers, learners, and employers concerned with the attainment of DD&E competencies, and finally, developers of competence-based programs in other professional areas.

Derivation and Classification

Robert Gagné (1970, 1972) has strongly influenced our thinking on the problems of levels of specification in the derivation of educational goals and objectives. Gagné distinguishes two major lines of derivation which he labels societal and educational, and nine levels ranging from national goals to instructional objectives. One of Gagné's points is that each line of derivation tends to skip a level, the societal derivation typically skips the level concerned with human functions (e.g., social communication) and the educational derivation typically omits the level concerned with manpower statistics relating to the relative numbers of jobs and roles required. In reviewing the data bases for educational R&D personnel, we found that both of Gagné's predictions about omissions were true and that they do have practical implications for the derivation of competencies.

One task we have completed is to examine the available knowledge base in terms of Gagné's nine levels of specificity and two sources of derivation. Briefly, our conclusions are these. Level 1 (National Goals or Man-in-Society Goals) and Level 2 (Social System Goals or Life Segment Goals) typically have been ignored or assumed in the derivation of educational RDD&E goals and objectives. And, at least until recently, our derivations have not dealt with Level 3 (Manpower Goals), with the consequences Gagné identifies, namely, that the relative frequencies of requirements are not considered. Currently, we have some information, although quite imprecise, on which to base quantitative estimates of personnel requirements.

Level 4, Human Functions, has been largely ignored in the social derivations, with the result that the general functioning of R&D professionals as persons in society or even as members of a work team have been given little consideration.

Level 5, Human Activities (e.g., following directions in completing applications), appears to be the typical entry point in recent efforts to define R&D requirements by analysis of observation, questionnaire or interview data. These data are summarized in frequency counts and cross tabulations. Patterns among activities may then be sought through correlational and factor analysis. From these data, we may in turn derive curriculum goals, broad objectives, and finally, instructional objectives. But by the time we have descended to these lowest and most specific levels, we discover that our linkages of the data base are quite tenuous.

After making an extensive review of the two largest and most complete data sources available to us--the Oregon Studies (Schalock, et al., 1972) and the AERA Task Force Studies (Worthen, et al., 1970, 1971)--we have concluded that each is useful but neither is very satisfactory for our needs. In our opinion, the 69 task categories and 226 competence categories developed by the AERA Task Force through interviews with over one hundred highly knowledgeable persons selected from a broad representative sample of educational RDD&E settings, represent the single most useful and comprehensive point of departure available to us. However, the AERA data provides only frequency of mention, correlational, and factor analysis data on briefly defined statements. But these data do present a fairly clear picture of overall patterns of RDD&E tasks and competencies.

For more detail, we have turned to the Oregon Studies and to our own Far West Training Consortium (Hood, et al., 1970) data. Unfortunately, the Oregon data, although rich and voluminous, have a maddeningly elusive quality. Perhaps their greatest value is the wealth of detail about the history, staff structure,

and political-institutional and intellectual contexts of educational DD&E projects. When one reads the seven (by my count, nine) development project case studies, much of the comfortable statistical certainty of the AERA factor analysis data dissolves in a sea of contradictions about what a development project is, how it should be staffed, what skills the personnel need, etc.

It is perhaps a sad commentary, either on our inability to produce or or inability to locate better data, but we have finally turned primarily to our own DD&E Consortium Task Inventory to provide more specific detail. This inventory was completed by 40 persons working in 32 educational DD&E projects in 1970. Their responses indicate the frequency of task performance, who performed what tasks, the judged criticality of the task, and whether the respondent learned the task on-the-job or in school. To the extent that we can say that our catalogue has an a priori data base, it will be dependent on the AERA and the Far West data. The AERA tasks and competencies will help to locate our development focus within the larger field of RDD&E. The Far West task data provides estimates of frequency and criticality cross tabled by levels of professionalization.

Currently, the map of our catalogue is represented by Figures 4, 5 and 6.

Figure 4

DD&E Competence Catalogue Table of Contents

I. Introduction

°Purpose

°How to use

II. The Data Base

°How derived

°How validated

°List of AERA/FWL data

III. The Competency List (Curriculum Objectives)

°Introduction/How Organized/How to use

°The Competency List

IV. Example Instructional Objectives and Test Item. (Confined primarily to Development of Competencies at entry-professional (M.A.) level derived from Development Series Competence Battery.

V. Technical Description

°Derivation and Classification

°Assessment and Critique of Approach

°Uses and Limitations

Figure 5

Sample Entry from Data Base Chapter

% Citing # Item

15% (T30) DEVELOPING NON-TEXTUAL MATERIALS

14% (C507) Knowledge of design stages in developing audio-visual or multi-media materials.

Competence References

37, 38, 53 (A3) Specify requirements for materials.

	Frequency			Criticality				
	Often	Some	Never	Essential	Desirable	Not Essential	Where Learned: Ojt	Where Learned: Sch
Phd	4	(7)	0	(9)	2	0	1	(4)
MA	(6)	(6)	2	(12)	1	0	(7)	2
BA	1	(3)	(3)	(4)	2	2	(3)	1

***

39, 42 (A5) Identify instructional materials.

	O	S	N	E	D	NE	Ojt	Sch
Phd	(4)	3	(4)	(6)	2	1	(4)	0
MA	(5)	3	(5)	(8)	1	1	3	(4)
BA	(6)	2	3	(6)	2	0	(4)	2

***

71-75, 83, 85, 95 (A21) Plan Evaluation of prototype.

	O	S	N	E	D	NE	Ojt	Sch
Phd	(6)	(6)	0	(11)	2	0	3	(6)
MA	4	(5)	2	(8)	2	1	1	(6)
BA	3	3	3	4	(7)	0	2	(3)

***

(Sample from Data Base Chapter)

Figure 6

Single Entry from the Competency List

38. Given a problem statement (including definition of need, general objective, approach, constraints and resources) write a set of development specifications for: (See also statements 37, 53).

(a) a short chapter in a conventional textbook. (See Data Section T23-C376).

(b) the same content as 38 (a), but for a linear programmed text (See Data Section T28-C376).

(c) the same content as 38 (a), but for a slide/tape program (See Data Section T30-C507).

(Sample for Competence List)

CONCLUSIONS

What have we learned thus far?

1. We are working with a weak data base to derive DD&E competence statements. None of the major studies--AERA, Oregon or Far West involve much more than 40 or 50 subjects with actual, extensive experience in development. Hence our samples are small and may not be representative. Moreover, the information obtained is not sufficiently detailed to be of much use in phrasing specific instructional objectives.
2. At least in educational development, there is not yet a well-defined convergence of opinion among experts about what development is or what competencies are needed for specific levels of professionalization. We anticipate that our DD&E catalogue will be heavily criticized. Part of this may be well deserved but part of it will be due simply to the fact that different "experts" will not be able to agree on any specific approach. Perhaps the greatest value the catalogue may serve is as a compendium of what we know and as a stimulant to better define what we need.
3. Self ratings do provide a useful, inexpensive way to "survey" broad areas of competence. The data we have obtained thus far indicate that students entering our DD&E training program do so with markedly different profiles of competencies which justify the tailoring of individual program objectives to their needs.

4. Development of adequate knowledge and job sample tests have proven to be difficult and expensive.
5. We have some, but inadequate evidence that we will be able to empirically validate at least the concurrent validity of our assessment instruments.
6. Our decision-oriented, designed-for-the-user approach has been well received. Pilot tests suggest that the diagnostic phase will work. It will take several more months of development followed by operational testing before we can safely make any predictions about how the assessment battery works.
7. A Bayesian approach to decision making, which is capable of taking a broader set of objective data and subjective estimates into account in classifying students with respect to their proficiency seems quite feasible. Again, we have yet to test the methodology in the field before we can be assured that users will actually employ the methodology.
8. Although we have made a beginning, we are a long way from professional certification based on inexpensive, validated assessment.

REFERENCES

Brown, R. V. Research and the Credibility of Estimates. Boston, Mass: Division of Research, Harvard University Business School,1969.

Campbell, D. T., & Fiske, D. W. Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. Psychological Bulletin, 56 81-105.

Clark, D. L., and Hopkins, J. E. A Report on Educational Research, Development, and Diffusion Manpower, 1964-1974. Bloomington, Indiana: Indiana University Research Foundation, 1969.

Elam, S. Performance-Based Education: What Is the State of the Art? Washington, D. C.: American Association of Colleges for Teacher Education, 1971.

Ferguson, R. L, & Novick, M. R. Implementation of a Bayesian System for Decision Analysis in a Program of Individually Prescribed Instruction. Iowa City, Iowa: ACT Publications, (P. O. Box 168), 1974.

Gagné, R. M. The Conditions of Learning. (2nd ed.) New York: Holt, Rinehart & Winston, 1970.

Gagné, R. M. Taxonomic Problems of Educational Systems (Presentation to NATO Conference on Utilization of Human Resources, Lisbon, Portugal, 11-15 June, 1973). Gainesville, Florida: Florida State University, 1973.

Hood, et al. "Design of a Functional Competence Training Program for Development, Dissemination, and Evaluation Personnel at Professional and Para-Professional Levels in Education." Berkeley, Ca.: Far West Laboratory for Educational Research and Development, December, 1970.

Merwin, J. C. Performance-Based Teacher Education: Some Measurement and Decision-Making Considerations. Washington, D. C.: American Association of Colleges for Teacher Education, 1973.

Popham, J. W., Cary, G. R., Chilstrom, B., Leps, A. L., Miller, T. M., and Saxe, R. "Performance Tests for Instructional Developers: An Exploratory Investigation." Educational Technology, March, 1974 pp. 57-61.

Schalock, H. D., and Thomas, G. P. Some Observations on the Contribution of the Oregon Studies in Educational RDD&E to the Development of Competency Based Personnel Development Programs. Monmouth, Oregon: Teaching Research, 1973.

Schalock, et al. The Oregon Studies in Educational Research, Development, Diffusion, and Evaluation, Volume I, Summary Report. Monmouth, Oregon: Teaching Research, 1972.

Turner, Richard Levels of Criteria, Appendix A, in The Power of Competency-Based Teacher Education, (ed.) Benjamin Rosner and others. Princeton, N. J.: Educational Testing Service, Washington, D. C.: U. S. Dept. of Health, Education and Welfare, Office of Education, 1971.

Worthen, B. R., Millman, J., Page, E. B., & Brezezinski, E. J. Development of a Pilot Test of Selected Competences in Educational Research, Evaluation, Development and Diffusion. (AERA Task Force Technical Paper No. 29). Washington, D. C.: American Educational Research Association, 1971.