DOCUMENT RESUME

ED 095 225 95 TM 003 965

AUTHOR Dunn, James A.; And Others
TITLE Effectiveness of Educational Laboratories and Centers: A System For the Evaluation of Educational Research and Development Products. Final Report.
INSTITUTION American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO AIR-21800-11-72-FR
PUB DATE Feb 73
CONTRACT OEC-0-70-4891
NOTE 292p.

EDRS PRICE MF-$0.75 HC-$13.80 PLUS POSTAGE
DESCRIPTORS Classification; Educational Research; *Evaluation Criteria; *Evaluation Methods; Evaluation Techniques; *Guides; Rating Scales; Records (Forms); *Research and Development Centers; *Summative Evaluation; Taxonomy

ABSTRACT

A system was designed and pilot tested for the evaluation of the products of educational research and development centers and laboratories. The products developed were: (1) a detailed specification of the evaluation procedure; (2) an empirically derived, validated, and reliable product taxonomy; (3) criteria for evaluation; and (4) the forms, instructions, manuals, and guidebooks necessary for product inventory, classification, evaluation, data tabulation, and summarization, and reporting of results. During field testing the first large-scale inventory and description of laboratory and center products ever made was carried out. A hitherto undeveloped theoretical model, based on the psychometric "method of successive judgments," was identified, elaborated, and operationalized in a new rating scale format. A 10 percent sample of completed products was selected on which to try out the evaluation system. Half of the products were evaluated with the experimental successive judgments rating method; the other half, with the usual single judgment method. Comparisons of the rating methods, the results of the product evaluations, suggested revisions in the evaluation paradigm and materials, and cost projections for operation of the system in alternative administrative contexts, were given. (Author/RC)

FINAL REPORT

Contract No. OEC-0-70-4891

EFFECTIVENESS OF EDUCATIONAL LABORATORIES AND CENTERS: A SYSTEM FOR THE EVALUATION OF EDUCATIONAL RESEARCH AND DEVELOPMENT PRODUCTS

James A. Dunn
Lauri M. Steel
Judith M. Melnotte
Ann H. McCormick

American Institutes for Research
Palo Alto, California

February 1973

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education
National Center for Educational Research and Development
and the
Office of Program Planning and Evaluation

ABSTRACT

The purpose of this project was to design and pilot test a system for the evaluation of the products of educational research and development centers and laboratories.

The products developed were: 1) a detailed specification of the evaluation procedure; 2) an empirically derived, validated and reliable, product taxonomy; 3) criteria for evaluation; and 4) the forms, instructions, manuals and guidebooks necessary for product inventory, classification, evaluation, data tabulation and summarization, and reporting of results.

During field testing, the first large-scale inventory and description of laboratory and center products ever made was carried out. Over 3,800 pages of product information were collected in this effort.

Regarding product evaluation, a hitherto undeveloped theoretical model, based on the psychometric "method of successive judgments," was identified, elaborated, and operationalized in a new rating scale format.

A 10% sample of completed products was selected on which to try out the evaluation system. Half of the products were evaluated with the experimental successive judgments rating method; the other half with the usual single judgment method.

Comparisons of the rating methods, the results of the product evaluations, suggested revisions in the evaluation paradigm and materials, and cost projections for operation of the system in alternative administrative contexts, were given.

PREFACE

The evaluation procedures reported herein were developed by the American Institutes for Research for the U.S. Office of Education for use in assessing the products of educational research and development centers and laboratories. The guidelines for the development of this system were that the system should be:

- General enough that it can be used to evaluate a wide spectrum of educational research and development products.
- Simple enough that it can be operated with a minimum of staff support.
- Flexible enough to be implemented either by an internal governmental agency or externally by an independent contractor.
- Broad enough to serve possible expanded functions under NIE or USOE.

In developing the system, close contact was maintained with NCERD's network of university based Research and Development Centers and Regional Educational Laboratories. Numerous meetings were held with directors of the laboratories and centers, with representatives of the CEDAR Executive Committee, and with representatives of NCERD's Division of Research and Development Resources. In those meetings the evaluation paradigm, procedures, and materials used in the project were reviewed, discussed, and revised.

In addition to formal meetings with various sub-groups of laboratory and center directors, all laboratory and center directors were consulted at specific points in the system development process. Directors were asked to review and comment on the proposed evaluation criteria. They were asked to nominate and review candidates for the evaluation panels. And they were sent copies of the proposed evaluation materials for review and comment.

The resulting product evaluation system was pilot tested in May, 1972 by two separate, independent, groups of evaluators. Each group was comprised of subject matter specialists, product developers, evaluators, and product users. Both groups of evaluators independently critiqued the evaluation system after they completed their evaluations.

This report provides a detailed summary of the evaluation procedures, the results obtained from the pilot test, and recommendations for revision and future implementation of the system.

Special thanks are due to those laboratory and center directors and OE personnel who have been so helpful in this endeavor.

TABLE OF CONTENTS

LIST OF FIGURES

PART I

SYSTEM DESIGN

Chapter 1

INTRODUCTION

In 1963, the Research and Development Centers Program was established under provisions of the 1954 Cooperative Research Act, Public Law 83-531. An R&D Center was "conceived as a place where a critical mass of interdisciplinary talent and other resources could be focused on a significant educational problem" (USOE, 1969, page 75).[1] Between 1964 and 1967, ten research and development centers were established at major universities across the country.

In 1965, Title IV of the Elementary and Secondary Education Act, signed into law April 11, amended P.L. 83-531 to provide for the establishment of a series of independent, non-profit, regional educational laboratories. Their mission was to engage in educational research and development and to "speed the intelligent application and widespread utilization of the results of educational research and development" (USOE, 1969, page 71). Contracts for the first eleven laboratories were signed in February, 1966. By September, a total of twenty laboratories had been funded.

All told, during the three year period 1964-67, thirty laboratories and centers were established. In addition, two research and development centers focusing on vocational education, a National Laboratory for Early Childhood Education, with sites at six major universities, and two Educational Policy Research Centers, were also established.[2] During this period annual Federal funding for R&D efforts had increased more than 500% (Boyan, 1969).

---

1 Reference citations are listed in the Bibliography starting on page 145.

2 Within six years of their founding, approximately one-third of the agencies had been terminated. This is an amazingly short life span in view of the findings of Projects Hindsight and Traces that leadtimes of 30 years and of 9 years for the application of basic and applied research findings respectively are needed for general engineering problems. These findings pertained to the relatively well defined "hard sciences." Even greater leadtimes would, presumably, be needed for the less well systematized behavioral sciences and education.

Since their inception, through FY 1972, laboratory and center funding alone has totaled more than $180 million. This excludes building grants and all ancillary supplemental and collateral support funding received through sole source and other competitive grants and contracts.

ORIGINS OF THIS PROJECT

In the years immediately following the formation of the laboratory-center network, evaluation concerns were directed, of necessity, toward the assessment of the potential of various agencies for future contribution.

In 1966, the year that the laboratories and most of the centers were opened, the Commissioner of Education, Harold Howe, commissioned Francis Chase to undertake a special evaluation of the laboratories and centers, in order to obtain information and advice about the various agency operations. The Chase study (1968) took slightly more than two years to complete and was based on personal site visits and interviews.

It is important to remember, however, that Chase was not commissioned to evaluate laboratories and centers *per se* but rather to evaluate the potential that the laboratory and R&D center system had for eventual significant contribution to education. Chase, nevertheless, spent considerable time in his final report emphasizing the eventual importance of the evaluation of agency products and their impact. The present project is an effort to address one aspect of that recommendation.

PROJECT OBJECTIVES

The specific objectives of this project were: 1) to develop a procedure for the multi-dimensional evaluation of products issuing from laboratories and centers; 2) to pilot test that procedure using a small sample of laboratory and center products; and, 3) to suggest whatever revisions of the procedure seem appropriate based on pilot test results and the critical comments of consultants and product evaluators.

The aim of this project, then, was to provide a tested procedure for the systematic evaluation of federally supported R&D products irrespective of the organizational structure under which those products were developed.

Two types of products were to be considered: those products deriving from systematic developmental efforts and which often (although not necessarily) have some commerical value; and those deriving from basic and applied research efforts which result in the generation of new knowledge, i.e., in the expansion of the knowledge base on which new educational efforts might be based.[1]

ASSUMPTIONS AND LIMITATIONS

Initial attention was directed toward acquiring an understanding of the evaluation procedures utilized in the past, toward ascertaining current and projected assessment needs, and toward identifying the reality constraints that would be imposed upon the operation of the newly developed evaluation system should it be adopted.

Subsequently, attention shifted to the identification of the specific working assumptions, i.e., the "conditions" that would have to be met to assure reasonable system practicality. The more salient of those assumptions were:

1) The evaluation should be as objective as possible.
2) The unit of evaluation must be the product itself.
3) The procurement of products, and of all product supporting documentation to be used in the evaluation, should be through the product developer.
4) The final evaluation of a product should be based on the collective judgments of a panel of experts.
5) Product developers should participate in the identification of product evaluators.
6) The results of a product evaluation should be provided to the product developer as well as the funding agency.
7) Evaluators should have the opportunity to file minority reports if they so choose.

---

[1] More detailed definitions of "knowledge" and "developmental" products may be found in Appendix A in the instruction manuals for the completion of Product Reporting Forms.

8) The product developer should have the opportunity to file an evaluation rejoinder if he so chooses.

9) There should be provision for the re-assessment of products when conflicting results suggest it is appropriate.

Chapter 2

THE EVALUATION PARADIGM

In the course of designing the paradigm to be followed in product evaluation, four main theoretical prototypes were considered. They were

1) independent field test models,
2) independent field reader models,
3) developer self-evaluation models,
4) site visitor models.

Ten alternative procedural models are subsumed under these four basic categories.

CATEGORY 1: INDEPENDENT FIELD TEST MODELS

Independent field test evaluations are those evaluations which are based upon systematic, empirical evaluation by an independent agent. There are at least three main forms of such evaluation efforts.

Experimental Evaluation. In this variation the materials, products, plans, etc., to be evaluated are submitted to controlled, experimental study. Examples of this type are: Consumer's Union, the Underwriters' Laboratory, and replication studies as conducted by the American Chemical Society. The advantage is that they are impartially performed, potentially rigorous, empirical validations. The disadvantages are that such procedures are typically very expensive and time demanding.

Field Evaluations. In this model the product is already in field use and an evaluator is called in to examine the effectiveness of the products. Examples of this type were the national Head Start and Follow Through evaluation efforts.

As in the experimental validation effort, the major advantage of this type of evaluation is that judgments are made on empirical evidence of effectiveness. A major disadvantage, in addition to expense, is the lack of control by the evaluator of possible confounding factors. These difficulties range from lack of being able to establish adequate base lines (e.g., pre/post testing, control groups, etc.) to difficulty in ascertaining that the product was indeed implemented as intended.

User Evaluation with External Review. In this model, the user performs his own evaluation and then an independent evaluator is called in to assess the quality of that evaluation. Examples of this model are the Hawkridge (1968) studies of exemplary compensatory education projects and the independent assessor procedures used recently by OE. These procedures are quite inexpensive as far as independent evaluation is concerned and, as in the case of all good evaluations, are still based on empirical evidence. This procedure is dependent, however, on 1) identifying users conducting independent evaluations, and 2) the quality of user evaluation. As found in the Hawkridge studies, the frequency of high quality user evaluation is relatively low but those that are found to be of adequate design and execution are quite useful.

CATEGORY 2: INDEPENDENT FIELD READER MODELS

In this type of procedure evaluation is based on the judgment of experts pursuant to an in-depth analysis of the products to be evaluated. There are two basic types of field reader models; one where the readers serve as individual consultants, i.e., where their inputs are made separately, and the other where the field readers serve jointly as a group.

The formal aspects of independent versus group reader service are not as significant, aside from considerations of time, coordination, and cost, as are the conditions surrounding their evaluation efforts, i.e., whether they serve in essentially a passive judiciary role with a single, unilateral information input, or in an active, interrogatory role where there is reciprocal information exchange.

The Single Input Evaluation Model. In this type only one input of information is made to the evaluator. Examples of this type of evaluation are the AIR Creative Talent Award Program, OE proposal reviews and the like. This is often the model used to maintain equity of opportunity in competitive situations and to increase the possibility of inter-judge reliability. The single input evaluation lends itself very well to "blind" evaluation, it is simple to administer, and relatively low in cost. A very heavy burden is placed on initial data specification, however. Not only must all data needs be specified in advance, but those needs must be clearly indicated to the

data suppliers. The *a priori* identification of all data needs for new evaluation procedures is a major task, but one which may be approached empirically if several reiterations through the process are possible.

Information Loop Models. This type of evaluation model is an open information model where, in the event that an evaluator feels more information is necessary, it can be obtained; or in the event that an evaluator wishes to confirm a tentative conclusion with more data, he may do so. The major advantage of such a model is that it avoids the necessity of complete *a priori* specification of data on which judgments are made. The Information Loop Model is somewhat more expensive to conduct than the Single Input Evaluation Model. The expense tends to increase as the number of information loops increases. It is considerably dependent on the evaluator's initiative and, as such, may have low inter-judge reliability unless all data received through the various information loops are pooled before final judgments are made.

Both the single input and information loop models may or may not include a meeting of the independent evaluators in which they prepare a joint, summary evaluation based on their various independent judgments.

The overall advantage of field reader models is that they are considerably less expensive than field test models, yet they still encourage careful, detailed analysis of actual products. In addition, where empirical data are available (from whatever source: developer, user, or some other third party) they can be capitalized upon.

The overall disadvantages of the field reader paradigm are 1) difficulties of coordination, and 2) some products, such as very complex, not yet fully developed and "intangible" products (e.g., services) may not readily lend themselves to convenient packaging, communication by the mails or telephone, etc.

CATEGORY 3: SELF-EVALUATION MODELS

These are models in which the evaluation is conducted by the developer himself. They are of two types: unreviewed self-evaluation and self-evaluation with external review.

Unreviewed Self-Evaluation. This is the type of evaluation wherein an independent developer evaluates the product he himself has developed and does not formally subject his self-evaluation to external review. The methods, findings, and conclusions of the evaluation are unrefereed. This has been the traditional pattern for textbooks, scholarly works, and the like. In this model external evaluation is, of necessity, indirect. Some types of indirect evidence used in the past are the stature of the editor/publishing house agreeing to publish/distribute the work, and the extent of professional endorsement of the product or report.

Self-Evaluation with External Review. In this model the individual developer conducts his own evaluation of his product but submits the results of his evaluation (and the products) to external review. It is the counterpart of User Evaluation with External Review. However, one could reasonably suspect a higher degree of bias inasmuch as it is the developer himself conducting the review. This type of evaluation, however, does offer some opportunity for R&D product evaluation to be based on empirical evidence.

One of the practical disadvantages of too heavy a reliance on this type of information is that developers may have far less systematic empirical evidence regarding the effectiveness of their products than one would suppose. Evaluation during the course of product development is often conducted for its immediate practical value and hence is not written up and/or summarized in a form that is amenable to convenient communication to others.

CATEGORY 4: SITE VISITOR MODELS

The common element of the various visitor models is that a personal visit takes place. The purpose of the visit may range from simple data collection to fairly extensive interaction with the principals. Although it often occurs that way, the site visitor model does not necessarily imply that the visit be unstructured, that the marshalling and presentation of data be of the "show and tell" variety, nor that judgments need be based on simple opinion or impression. There are at least three basic forms of site visitor models.

The Developer Site Visit Model. This is perhaps the most frequently encountered model. A panel of experts visits a development site, sometimes

with only minimal preparation and little structure to the visit. One of the great advantages of this approach is that visits can be convened on relatively short notice and executed in a relatively brief period of time. They are also quite flexible and can be given a variety of charges quite easily. For success, however, visitors must be quite knowledgeable of the products they are to evaluate and very familiar with points of difficulty they might encounter.

Visitors cannot be expected to function well as evaluators if they receive only brief preparation, do not share common standards, and view products and issues from a widely disparate set of perspectives. In the absence of a clear cut structure for the site visit, evaluations tend to wander and installations being visited frequently resort to promotional presentations in order to impress the visitors.

Under ideal conditions, the site visitor comes well briefed as to the major purpose and mission of the agency, the products they have developed, and the criteria and standards by which the evaluation should be effected. The agency director, similarly, should be prepared to present detailed factual evidence regarding those criteria. Unfortunately, the brevity of most site visits frequently militates against such detailed presentations.

The User Site Visit Model. This is the second form of visitor-based evaluation. In this model, evaluators visit areas where the product is in use rather than where it was developed. This is analogous to User Evaluation with External Review, except instead of a review of an explicit user evaluation, informal interviews and observations by the visitors are substituted.

The Structured Visit Model. Still another form of the visitor model is the Structured Visit Model. In this procedure, a great deal of information regarding products, developer evaluation efforts, sponsor concerns, etc., is assembled and provided the evaluators well in advance of their visit to either a developer or user site. Much of this data may in fact have been pre-analyzed, and condensed by field readers, well in advance of the visit.

The advantage of a structured visit procedure is that incomplete prototype materials, very expensive or complex products, "soft" products such

as the research training contributions, and the consultation services of an agency may also figure in the evaluation.

GENERAL GUIDELINES

In addition to an analysis of the assumptions, advantages and disadvantages of the foregoing models, the following assumptions also played a role in the design of the specific operational paradigm to be developed and tested.

First, product evaluation should be predicated, to the extent possible, on primary data. The primary data for product evaluation should be the product itself, plus such support documents as rationale statements, needs analyses, and the like, produced by the developer. Field test and evaluation data are secondary data but may be especially useful if carefully evaluated as to quality before results are accepted.

Second, although it is the evaluator who uses data for making judgments, the evaluator need not be responsible for collecting the data. Such a requirement would result in inordinate demands on developers for data, and would in all likelihood, result in different evaluators using different data bases for the evaluation of the same products. It is also quite likely that data requested by a variety of evaluators at different times would not be as systematically marshalled as they might be for a single reporting.

Third, since during the first few applications of the evaluation model, the data collected may be incomplete or even erroneous due to poor definition, or misinterpretation, of the data requests, supplementary information might need to be collected. To insure data base constancy across all evaluations of a given product, any supplementary information obtained should be provided to all evaluators even though only one evaluator may have requested it.

Fourth, inasmuch as there are many products to evaluate, the evaluation of any set of products may be distributed across several months. This would facilitate the scheduling of evaluators, permit evaluators to participate in the

evaluation of more products, and, thus, tend to increase the number of highly desirable candidates who would accept the invitation to serve as evaluators.

Fifth, products should be evaluated only upon their "completion," i.e., when an agency is "through" with them, when it has carried them as far as their responsibility dictates.

Sixth, products should be evaluated only once unless a reappraisal is specifically requested by the developer.

Seventh, the most convenient location for the evaluation to take place is in the office of the evaluator. This would imply that all information necessary for the evaluation of the product, including a copy of the product itself, can be made available to the evaluator, presumably through the mails. This is clearly not possible for all products. Some products, such as mini-courses, are too expensive to make available to six to nine evaluators for several weeks each. Some products, such as IPI, are too complex to export physically and can only be "seen" in places where they have been installed.

After careful consideration of factors such as these, and the relative advantages and disadvantages of the various general procedural models discussed earlier, a tentative evaluation paradigm was constructed to meet the anticipated operational constraints imposed by government projections. This paradigm was then reviewed by a panel of consultants, OE staff, and laboratory and center directors, revised per consultant recommendations, and circulated by mail to all laboratory and center directors, in October, 1971, with a request for reactions, comments, and suggested revisions. It was this model that was then implemented in the pilot test.

PARADIGM SUMMARY

In summary, the paradigm followed in the pilot test involved several functionally discrete steps, each of which is described briefly below.

Step 1. Product Identification. The first step in product evaluation is the identification of the products to be evaluated. Because of the potential implications of product evaluation, laboratory or center directors, themselves, are considered to be the only appropriate source of information about products

issuing from their respective agencies. Thus, laboratory and center directors should specify those products from their agencies which are ready for evaluation, i.e., which are completed; describe the characteristics of those products and the contexts in which those products should be considered; and, should they wish to do so, provide any special factors or material, e.g., evaluation results, which they wish to have considered at the time of product evaluation. This information is obtained via Product Reporting Forms. Descriptions of these forms, and the results of the pilot test of this step are summarized in Chapter 5. Sample Product Reporting Forms and the instruction booklets for completing those forms are contained in Appendix A.

Step 2. Classification of Products for Evaluator Assignment. One of the assumptions underlying the design of the evaluation system was that products should be evaluated only by individuals who had technical-substantive expertise in the product area. Thus, products need to be classified according to their substantive domain. All products reported as ready for evaluation (i.e., "completed") need to be classified according to an empirically derived products classification. Chapter 6 summarizes the products' classification taxonomy and the results of the pilot test of this step.

Step 3. Selection and Training of Evaluators. Nominations of potential evaluators for the specific topic areas in which products are to be reviewed must be obtained. The resulting lists of nominees, one list for each product area, should then be submitted to agency directors for review and to the appropriate governmental offices for approval. Final selection of panel members for each product group is then made by the evaluation coordinator.

A central meeting of the evaluators should be held in which they can be introduced to the nature and purpose of the evaluation system and trained in the use of the evaluation instruments. At this time they may also be given all necessary product evaluation materials and other support materials.

The methods for this stage of the evaluation, and the results of the pilot test of this step are found in Chapter 7. The Evaluator's Manual and copies of the various Product Rating Forms are found in Appendix B.

Step 4. Product Procurement. The procurement of products for evaluation may run concurrently with Step 3. Upon identification of the products to be

evaluated, the respective laboratory and center directors should be notified and copies of the products requested. In addition, copies of the Product Reporting Forms for the products requested should be returned to the appropriate directors who are asked to confirm the information contained therein, or to revise or update it as they see fit. This is to insure that the product director has yet another opportunity to make substantive input to the evaluation of his product, and to verify the data base that would be used in the evaluation of that product.

(All agencies were most cooperative. Their prompt assistance in supplying sample products did much to facilitate the pilot test. In most instances products were supplied on a loan basis. In some instances products were donated outright; in others product costs were borne by the evaluation coordinator.)

Instructions and recommendations with regard to product procurement may be found in Chapter 8.

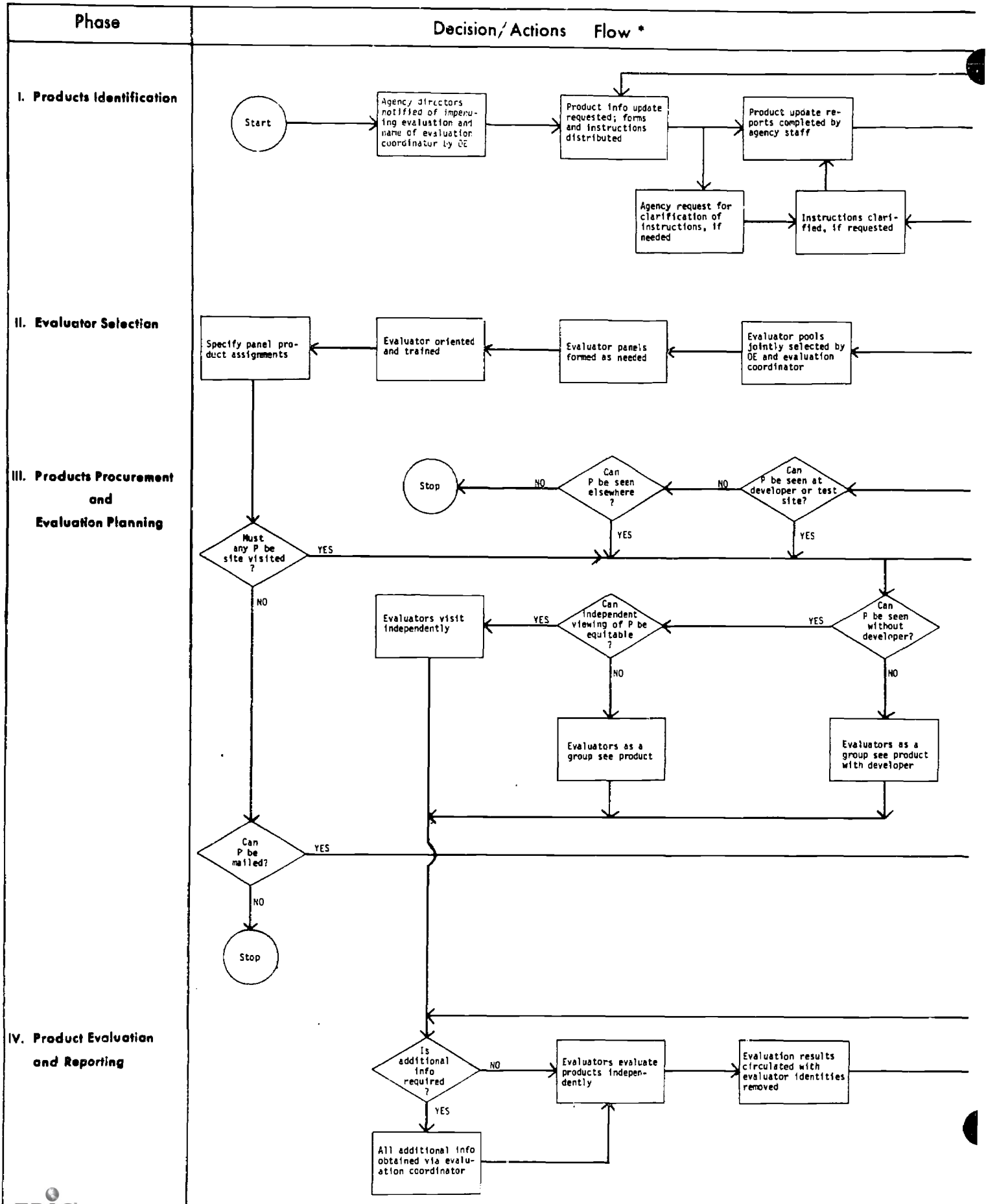
Step 5. Product Evaluation. Normally the majority of products are reviewed privately by evaluators in their own home offices. In those cases where it is not feasible to send the product to each evaluator, the evaluation coordinator will devise alternative arrangements. In one instance evaluators may need to review a product at a local operating site; in another it may be necessary to arrange for all evaluators to review the product at a central location.

After initial independent product judgments are made by each of the evaluators, the results should be circulated, along with supporting arguments but without rater identification, among all panel members. The evaluators are then asked to reconsider their initial judgments in light of the arguments presented anonymously by the other panel members. Following that, panelists are asked to either reaffirm or revise their initial judgments. Sample Rating Summary Sheets are provided in Appendix C.

Recommendations regarding the coordination of the evaluation effort are presented in Chapter 9. The results of the pilot test are presented in Chapter 10.

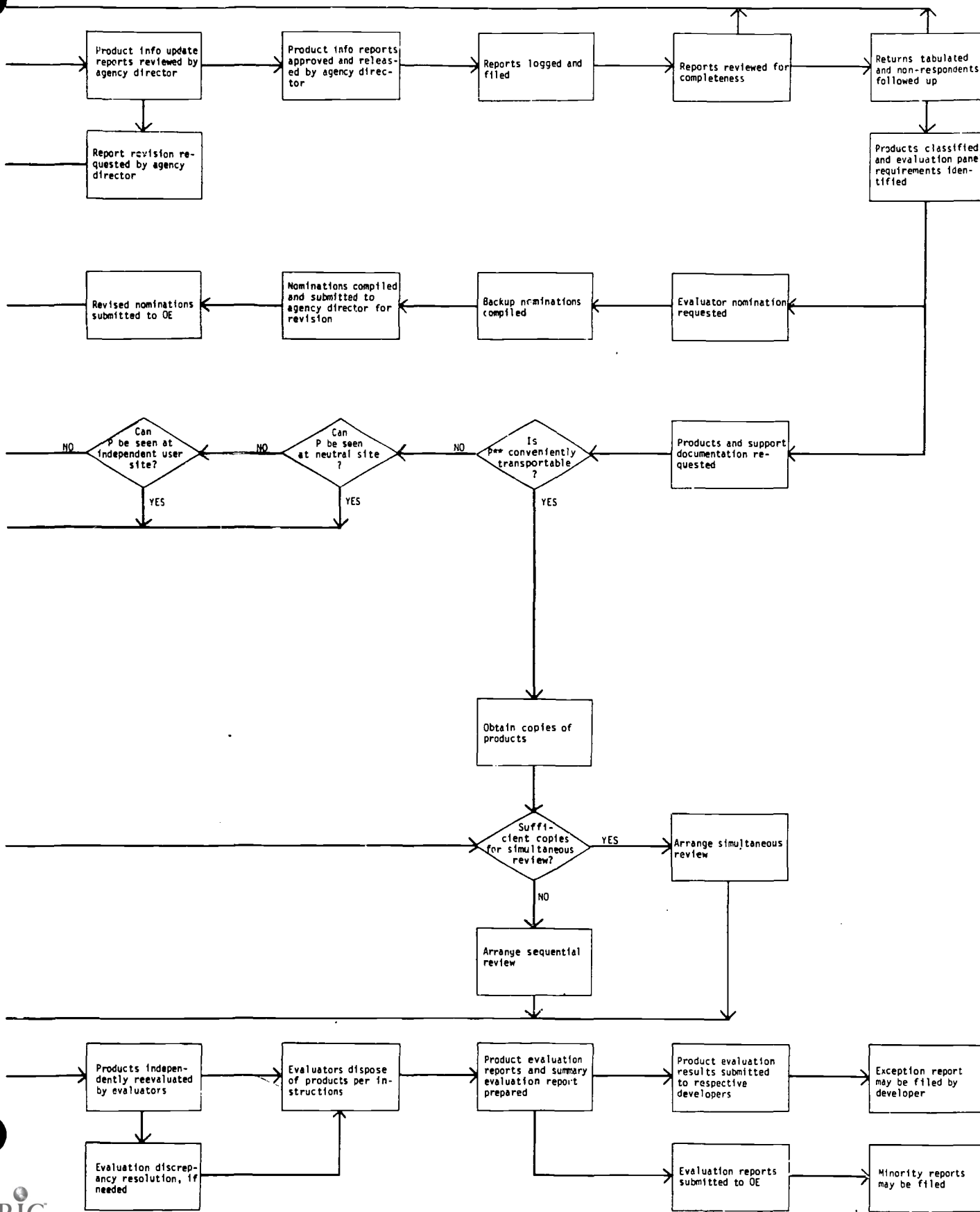
This paradigm is presented in greater detail in Figure 1 and in the pages that follow.

Figure 1 GENERALIZED EVALUATION PARADIGM



* Unless otherwise specified, all entries refer to actions or decisions by the evaluation coordinator.

** P = product


Product info update reports reviewed by agency director
Product info reports approved and released by agency director
Reports logged and filed
Reports reviewed for completeness
Returns tabulated and non-respondents followed up
Report revision requested by agency director
Products classified and evaluation panel requirements identified
Revised nominations submitted to OE
Nominations compiled and submitted to agency director for revision
Backup nominations compiled
Evaluator nomination requested
NO
Can P be seen at independent user site?
NO
Can P be seen at neutral site ?
NO
Is P conveniently transportable ?
Products and support documentation requested
YES
YES
YES
Obtain copies of products
Sufficient copies for simultaneous review?
YES
Arrange simultaneous review
NO
Arrange sequential review
Products independently reevaluated by evaluators
Evaluators dispose of products per instructions
Product evaluation reports and summary evaluation report prepared
Product evaluation results submitted to respective developers
Exception report may be filed by developer
Evaluation discrepancy resolution, if needed
Evaluation reports submitted to OE
Minority reports may be filed

THE DETAILED PARADIGM

The specific steps in the decisions/actions flow of the evaluation system are:

1) The appropriate OE administrator sends letters to the directors of the laboratories and centers notifying them of the pending evaluation and designating the evaluation coordinator.

2) The evaluation coordinator sends an overview of the evaluation procedures to all directors and alerts them that several man-days will soon be required to fill out or update product reporting forms and to assemble and transmit products and necessary support documents.

3) The evaluation coordinator sends product reporting forms and instructions to the laboratories and centers. The product reporting form contains questions regarding the description and nature of a product; e.g., objectives, target audience, effectiveness as indicated by data, etc.

4) Agency staff complete the forms. If a question arises regarding the completion or submission of the form, the respondent calls the evaluation coordinator for clarification. Upon completion of the form, the agency director reviews the report and approves it for release to the evaluation coordinator. If the product reporting form does not meet the director's approval, he recycles it through his agency.

5) The coordinator receives the form and checks it to make sure all information is complete. This task includes verifying that all forms have been received, that no known product has been omitted, and that all forms have been correctly and completely filled out. Should the missing information be minor, it is collected by telephone. If it is extensive, the form is returned with a request to supply the needed information.

6) The evaluation coordinator then tabulates receipts and all non-respondents are followed up. The first follow-up is made by mail two weeks after the report due date. The second is made by telephone four weeks after due date. Agencies not responding within six weeks of the due date are referred to OE for follow-up.

7) The evaluation coordinator uses the product reporting forms to organize products by topic area and to identify the number and types of evaluation panels to be required. A topic area will typically contain eight to ten products.

   Notice that topic areas are defined before the evaluators are selected. In this way the specific skills and experience which the evaluators must have are identified before evaluators are solicited.

   Products classified under one of the existing product categories will be evaluated by the appropriate existing panel. Products not appropriate for evaluation by one of the regularly nominated panels will be accrued until there are sufficient number of similar products to

warrant forming a new panel by the procedures above. Sufficient numbers of products to warrant panel formation may accrue by combining low frequency categories if such a combination is conceptually meaningful.

8) The evaluation coordinator solicits nominations for product evaluators from: the Past President, President, Vice Presidents, and President-Elect of AERA; the presidents and executive committees of APA Divisions 15 and 16 and of other appropriate national professional associations; the directors of the laboratories and centers; and from appropriate governmental agencies.

   Nominations are made for specific topic areas.

   If necessary, backup nominations are also made by the evaluation coordinator. Backup nominations may be drawn from such sources as Senior Fellows of professional organizations and editorial boards of professional journals.

9) The evaluation coordinator submits the list of nominees for each area to the laboratory and center directors for their review, addition, and/or deletion; he updates the list of nominees per feedback from directors and submits the lists to OE for final approval.

10) Upon receipt of the approved evaluator lists, the coordinator queries evaluators as to their willingness to serve and the times and extent to which they will be available.

11) The evaluation coordinator designates, from the approved list, panels of evaluators for each of the groups of products to be evaluated.

   The criteria for the selection of evaluators are:

   a) Evaluators must be known and respected in their fields.

   b) Evaluators serving as subject matter specialists should have substantive expertise in the topic area under consideration.

   c) Evaluation panel members must not all reflect the same theoretical bias.

   An evaluator will be asked to disqualify himself if:

   a) He has previously worked or consulted extensively on the product he is to evaluate.

   b) He has a vested interest, either financial or theoretical, in the product he is assigned to evaluate.

   c) If the product he is assigned to evaluate may be considered in direct competition with a product the evaluator has a vested interest in.

   d) The evaluator knows of any other reason to warrant his disqualification.

12) An evaluation panel for any given product area will consist of six to nine members. Specialists in the content area of the product shall predominate. However, there shall be at least one evaluation specialist and one consumer representative on each panel. Ideally, panel members should be able to serve for an extended period of time, i.e., for several evaluation cycles (for the evaluation of 20-30 products). No one should be appointed to a panel who does not expect to complete at least one full cycle. Panels may be updated by the evaluation coordinator as needed, however, from the list of approved evaluators for that area. The evaluator pools, i.e., the list of approved evaluators for the various content areas, will be reconstituted via the nomination and review procedure every two years.

13) Laboratory and center directors are notified of the products selected for evaluation, copies of the products and all relevant supporting documents are requested, and confirmation of the information on the agencies' product reporting form for each product is requested.

Usually ten copies of a product will be procured so they may be reviewed concurrently by the evaluators.

Occasionally, with expensive products, or products in limited supply, only one copy of the product will be procured and rotated among evaluators.

Occasionally, the coordinator may have to deal directly with publishers or distributors to obtain a product.

In cases where a product is too bulky or inconvenient to mail, the coordinator will determine an alternate procedure based on the specific circumstances. The evaluation coordinator's office may be used as an evaluation site. Evaluators may view the product individually at its site. If more than one site is available, each evaluator may travel to the most convenient site. Should it be necessary for all evaluators to view the product together, the visit will be arranged and monitored by the coordinator.

Because most products will be mailed, the evaluators will not have an opportunity to discuss their individual evaluations with each other. When joint site visits are necessary, opportunity for discussion will arise but should be actively resisted. This will tend to keep evaluation procedures consistent for all products.

14) Evaluators meet for an orientation-training conference. This should be a full day meeting. During this meeting evaluators are oriented to the evaluation procedure, review the criteria to be used, and execute several practice evaluations. After that, products not convenient for mail distribution are evaluated. Products amenable to mail distribution, or which require special field visits, will be evaluation subsequently.

15) After carefully studying the product, the evaluator makes his initial evaluation and completes the evaluation form. The criteria for judging the products are summarized in the Evaluators' Manual. Product ratings will be recorded on a series of rating scales. In addition to numerical

ratings on each of the criteria, the evaluator may also make written comments. The evaluator should be encouraged to elaborate on the frame of reference he is using when he makes his evaluation.

16) While evaluating a product, should an evaluator seek further information on it, he will request it of the coordinator, who will obtain the information from the appropriate agency and then inform all evaluators working on the product in question. This procedure will help assure that all evaluators work with the same information on any given product. This will also allow the coordinator to record the kinds of information that are requested so that forms, instructions, and procedures may be improved for the next cycle of evaluation, presumably the following year.

17) After evaluators have made their initial evaluations and submitted their independent reports and comments to the coordinator, the results will be circulated within the panel but without rater identities. Panelists will then be requested to reconsider the products in light of the judgments of the other panelists and to confirm or modify their original judgments, as they see fit.

18) Evaluators reconsider the products, complete their evaluations, and submit their final independent reports to the evaluation coordinator.

If there is more than a one-point discrepancy in the judgments of more than two evaluators, the discrepancy will be discussed jointly by the panel. If the discrepancy is resolved, evaluators may have a second opportunity to revise their judgments; otherwise, the variance, and its reasons, will be identified in the final report.

19) The evaluators will keep or return the products as instructed by the coordinator. Free products may be kept. Other products will be returned to the evaluation manager, to the appropriate agency director, or disposed of according to the instructions of the evaluation coordinator.

20) The evaluation coordinator will summarize and analyze the product evaluations. As a minimum, for each product the individual evaluator ratings on each criterion will be combined, through averaging, to form a summary panel evaluation. Instances of considerable disparity in judgment on particular criteria will be identified. The panel judgments for each of the criteria will then be plotted to yield an evaluation profile for each product.

21) Additional data analyses, such as those suggested in the following section, could also be completed at this time.

22) The evaluation coordinator submits the completed products file, evaluations, and evaluation analyses to the government. The names and backgrounds of the individuals comprising each evaluation panel will, of course, be reported. The judgments of specific individuals will not be reported, however.

23) Panel members may file minority reports if they wish.

24) A summary of the relevant product evaluations is sent to the appropriate laboratories and centers.

If a director has serious disagreement with an evaluation, he may request a re-evaluation. Such a re-evaluation would be processed by the evaluation coordinator with a different evaluation panel, but at the requesting agency's expense.

This re-evaluation option increases the system's capability to deal with unusual or extreme cases and would allow a laboratory or center to prepare a better case for its product.

These evaluation activities could be massed or distributed over time, depending on the needs of the government and the backlog of products to be evaluated. The larger the number of products to be evaluated during a given time period, the greater the problems of coordination. Once the backlog of accumulated products has been evaluated, however, the system could operate routinely as products are completed.

QUESTIONS THAT MAY BE ASKED OF THE SYSTEM DATA BASE

Given implementation of this paradigm, a number of very interesting, and potentially very crucial, questions could then be asked of the data base. For example:

1) How significant are the products produced by the various laboratories and centers?

2) How original and creative have their products been? What is the ratio of original products to all products?

3) How reasonable, in terms of cost and marketability, have the products been?

4) How effective are those products? How many products do, in fact, have effectiveness data?

5) What is the likely potential impact of those products?, on whom?, and in what areas?

6) Is there a difference in the work areas and outputs of laboratories and centers?

7) What proportion of output has been picked up and is being promoted by commercial interest?

8) Who are the primary publishers of laboratory and center products? Are they key publishers in their area? Is there broad representation across publishers?

9) What is the relationship of estimated impact of a product, the originality of a product, and the problem area it addresses?

10) What is the source of the most original, effective, and economically feasible products?

11) What is a reasonable base rate of productivity? Which agencies seem to be the most effective in product development?

12) Does targeted research for broad target populations have the same degree of quality and effectiveness as products for which there are more limited targets?

13) What is the character and form of the products? Is there variation in the form of solutions proposed, or do the majority of products tend toward a single approach, e.g., paper and pencil curriculum materials?

14) Given additional information regarding organizational structure, staffing patterns, management characteristics, etc., what, if any, relationship exists between organizational/structural variables and the types of problems various agencies select to work on, the significance and quality of their products, the practicality of their products, their overall level of productivity, the effectiveness of the products they produce, the overall level of originality and creativity they have contributed, and so forth?

15) What are the underlying characteristics, if any, that the highly effective agencies have that the minimally effective agencies do not have?

Many of these questions can be answered with data already in hand; others would, of necessity, require the accumulation of data resulting from the actual implementation of the evaluation system.

Chapter 3

THE CRITERIA

Upon initiation of the project, project staff began the accumulation of a large number of potential criterion items. There was often considerable overlap in many of the items collected and also considerable heterogeneity in their applicability across various forms of laboratory and center products.

As the criteria from the criterion pool were applied to various sample products, those that had overly narrow applications, i.e., those that could be used with only a few product types, were discarded. Similarly, those reflecting a high degree of redundancy were collapsed into larger, more general, criteria.

The goal was to select three to four criteria for each of four criterion groups: significance, quality, effectiveness, and practicality. Separate, though highly similar, criteria were used for knowledge products.

The criteria finally selected for use in the pilot study are summarized in Figures 2 and 3 and are described in detail in subsequent pages.

CRITERIA FOR THE EVALUATION OF DEVELOPMENTAL PRODUCTS[1]

Importance of General Problem. A problem is a recognized discrepancy between an existing state in education and a desired end state. As such, it may be described as an "educational need." In considering the importance of a problem, the question is "how crucial is it?" The magnitude of importance is a function of the number of people it affects and the intensity with which it affects them. A problem which intensely affects a large number of people is, of course, easily recognizable as an important problem. A problem that affects relatively few people, and only slightly, is easily recognized as being of little importance.

The difficulty of judging the magnitude of a problem's importance comes when judgments have to be made with regard to products affecting only a few

[1] Slight revision in the titling of three criteria has been made since the pilot test to improve clarity. See Chapter 11 for suggestions as to criterion reduction.

Figure 2

EVALUATION CRITERIA:

DEVELOPMENTAL PRODUCTS

IMPORTANCE OF GENERAL PROBLEM:	. . . degree to which problem is crucial to education . . . magnitude of the problem
RELEVANCE OF PRODUCT TO GENERAL PROBLEM:	. . . degree to which product clearly and directly relates to stated problem
COMPREHENSIVENESS OF THE PRODUCT AS PROBLEM SOLUTION:	. . . degree to which product meets the whole problem
CONTENT ACCURACY:	. . . informationally correct . . . a precise accounting and presentation
CONTENT CLARITY:	. . . an easily understood exposition . . . full, unambiguous explanations and directions
EFFECTIVENESS:	. . . degree to which product solves the problem . . . degree to which product meets its objectives
REASONABLE COST TO ADOPT/ IMPLEMENT, GIVEN OUTCOME:	. . . degree to which product is worth buying, given what might or will come of its use
REASONABLE COST TO USE/ OPERATE, GIVEN OUTCOME:	. . . degree to which product is worth continuing to use
SCOPE OF POSSIBLE MARKET:	. . . possible number of users, buyers, clients
AMENABILITY TO MARKETING:	. . . attractiveness of product . . . ease of acquisition and use
POTENTIAL IMPACT:	. . . likelihood of effecting change in educational practices, given all factors

Figure 3

EVALUATION CRITERIA:

KNOWLEDGE PRODUCTS

IMPORTANCE OF GENERAL PROBLEM:	. . . degree to which problem is crucial to education . . . magnitude of the problem
RELEVANCE OF PRODUCT TO GENERAL PROBLEM:	. . . degree to which product clearly and directly relates to stated problem
COMPREHENSIVENESS OF THE PRODUCT AS PROBLEM SOLUTION:	. . . degree to which product meets the whole problem
ORIGINALITY OF PRODUCT:	. . . extent to which product represents a unique contribution
QUALITY OF LITERATURE DISCUSSION:	. . . exhibits an awareness of current "state of the art" . . . appropriate to problem area
ADEQUACY OF RESEARCH DESIGN:	. . . appropriateness of statistical treatments . . . representativeness of sample
APPROPRIATENESS OF INTERPRETATION:	. . . justified by the data
REASONABLENESS OF CONCLUSIONS/ RECOMMENDATIONS:	. . . generally logical . . . substantiated by the findings
CLARITY OF PRESENTATION:	. . . an easily understood exposition . . . full, unambiguous discussion
POTENTIAL IMPACT:	. . . likelihood of effecting change in educational practices, given all factors

persons, but relatively intensely, as in the case of some special education programs. Difficulties may also be encountered with products that affect a larger number of people, but only modestly. It is at this point that the judgment of a problem's importance is most apt to be tempered by one's philosophy, experience, and professional commitment.

Relevance of Product to General Problem. Relevance refers to the degree to which the product under consideration clearly and directly relates to the stated educational problem. The product that is addressed directly to the heart of the problem has greater relevance than the product which deals only with some tangential aspect of the problem. For example, if the product developer indicates that his product is intended to help solve the problem of chronic poor reading in minority group children, a teacher's manual enhancing the story-telling abilities of primary grade pupils would be judged less relevant to the problem than a manual telling the teacher how to manipulate reinforcement techniques during reading instruction. This is not to say that the former product is not related to the teaching of reading; indeed, there are many who feel that verbal language ability is a necessary prerequisite to the enhancement of reading achievement. The product simply is not central to the problem as it was stated.

Comprehensiveness of the Product as Problem Solution. The comprehensiveness of a product depends on the degree to which the product meets the entire problem. If a product addresses all of the major facets of a problem, no matter how small or trivial the problem, then the product should be judged comprehensive. On the other hand, a product which deals with only a small portion of the general problem must be viewed as less comprehensive, regardless of the size of the effort devoted to the development of the product. It is not the size of the problem addressed which defines comprehensiveness; nor is it the size of the effort undertaken in the development of the product that counts. It is, rather, the extent to which the product addresses the whole problem, as it was stated on the product report form.

Content Accuracy. Accuracy refers to the extent to which facts, calculations, data, concepts, etc. presented in the product are informationally correct.

Content Clarity. Clarity refers to the extent to which the product text and/or materials are clear in their message. The materials should be easily read and understood. Directions for their use should be simple and straightforward. The user, whether he be student, teacher, administrator, etc., should not have to spend inordinate amounts of time trying to comprehend what is in the materials, the purpose of their existence, or how to use them.

Effectiveness. A product is effective to the extent that it works, i.e., to the extent that it meets its intended objectives.

The product *per se* typically does not include information on its effectiveness. The evaluator normally must base his judgment of the product's effectiveness on an examination of the reports and support documents submitted by the developing agency.

If an evaluator has information or knowledge about the effectiveness of the product under consideration, from sources other than those documents submitted in support of the product by the developing agency, that evaluator should notify the evaluation coordinator so that the additional evidence may also be made available to the other evaluators. Evaluators should be careful to avoid judging the effectiveness of a product on the basis of either opinion or prior judgment made as a consequence of evaluation results not currently supplied with the product, and, thus, not available to other evaluators. The judgment of product effectiveness must be based on a careful review of objective data.

If the product developer does not supply any evidence in support of his product's effectiveness, no judgment of product effectiveness can be made. The lack of any supporting evidence should be so indicated on the product evaluation form.

Reasonable Cost to Adopt/Implement Given Outcome. This criterion applies to what is commonly referred to as "purchase price." The question here is whether the product is worth purchasing given what it is expected to do. In some cases this question is fairly easy to answer. For example, a program which improves children's knowledge of classical music composers for $20 per pupil per year would probably be judged relatively expensive. On the other hand, some comparable expenditure, or even a considerably higher one, may be happily accepted if the outcome of the expenditure is highly valued. For example, it might cost many thousands of dollars to institute a new reading program. However, if it were effective in raising the reading level of non-readers to a level of independent reading competency, it might quite likely be judged worth the cost.

The main question here is not whether the cost of adoption is high or low, but whether the cost is reasonable, given what the product will do, i.e., whether the educational community is likely to get a good return for its investment.

Reasonable Cost to Use/Operate Given Outcome. This criterion is related to what is often called "operating costs." It applies to such routine ongoing expenses as replacement of consumable materials, equipment repair and servicing, periodic personnel costs, and the like. These are costs necessary for the continued use of a product after it has been acquired and installed.

The question here is once again not whether the costs for continued operation of the product are high or low, but rather, whether the expenditure of funds for continued operation is worthwhile, given the results accruing from product use.

Scope of Possible Market. This criterion refers to the product's theoretically possible market, not to its probable market, i.e., not to its estimated or projected sales. Here the emphasis is on what the potential size of the market could be if the product were effective and attractive, and clients could afford its purchase. In some discussions this criterion may also be referred to as the product's potential market.

While it is recognized that a number of qualifiers affect the realistic boundaries of potential markets, evaluators should nonetheless attempt to make a judgment about the possible scope of utilization of a product.

Some products, while very important, may be pertinent for only limited audiences. Thus, such products would have quite a limited potential market. Other products might have more general or pervasive application throughout all educational audiences. Products which contribute to solutions of more pervasive problems would have a wider potential market.

Amenability to Marketing. The question here is "Do you think the product, as it is presently formed, will lend itself to effective marketing?" That is, will someone be able to market it effectively? A number of factors enter into this decision: Is the product attractive? Is it assembled in such a way that it can be efficiently produced? Does it lend itself to convenient advertising, supply, classroom storage, etc.? In some discussions this criterion may also be referred to as potential marketability.

Potential Impact. In assessing potential impact, evaluators should ask to what extent the product has the potential for improving educational practice on a major scale. The basic question is to what extent the product is likely to effect a change in educational practice considering all the characteristics of the product and other factors which may influence its adoption and utilization.

CRITERIA FOR THE EVALUATION OF KNOWLEDGE PRODUCTS[1]

Importance of General Problem. This criterion is the same as for developmental products.

Relevance of Product to General Problem. This criterion is the same as for developmental products.

Comprehensiveness of the Product as Problem Solution. This criterion is the same as for developmental products.

---

[1] These were the criteria as used in the pilot test. See Chapter 11 for suggested revisions in this list.

Originality of Product. An original product is one which represents an imaginative or ingenious approach to solving the general problem to which the product is addressed.

The originality may be in problem conceptualization, methodology, or interpretation. The uniqueness of the document's ideas and/or methodology, of course, may only be judged within the evaluator's knowledge and experience.

Quality of Literature Discussion. This criterion is not applicable to some types of knowledge products. For many, however, a literature review provides a strong integrating context.

The desirability for comprehensiveness in literature reviews varies with the type of knowledge product. Products whose sole purpose is to review literature need be, of course, very comprehensive. Citations should include all the major efforts in an area and probably many of the lesser known efforts. In other types of knowledge products, however, the review may be much less comprehensive; thus, this criterion is not synonomous with extensiveness.

In all cases where a literature review is part of the product, it should a) be appropriate to the specific problem area; b) make explicit the relationship of previous research to the problem area cited; and c) point out how the additional new research accommodates or enhances the previous citations. In addition, the researcher should exhibit: a) an appreciation of the current "state of the art;" b) total familiarity with recent, pertinent literature; and c) an attempt to interpret, synthesize, and evaluate the relevant literature.

Adequacy of Research Design. This criterion applies to only that subset of knowledge products concerned with research. Like originality, the criterion of design adequacy includes a variety of considerations. Clearly all conceivable aspects of design cannot be considered in detail. The intent of this criterion is to allow for a rather general judgment to be made about the overall adequacy of a product's research design.

Basic consideration should include at least the following, however:

a) the degree to which the design is suited to the problem;

b) whether the design represents a rigorous test of the stated or implied hypotheses;

c) whether potential error has been reduced and threats to validity minimized through such procedures as:

   1) random assignment of subjects,

   2) statistical or experimental control of intervening variables,

   3) sufficient numbers of subjects,

4) dependent variable instruments of sufficient validity and reliability,

5) sampling which allows for justifiable generalizing, or

6) acknowledgment and satisfaction of statistical assumptions, and the like.

Appropriateness of Interpretation. Appropriateness of interpretation deals with the degree of reasonable accord between the factual results of a study and the statements made about those results. The key issue is the degree to which interpretations or statements about the results are, in fact, justified by the data. Evaluators should be alert to misinterpretations, inappropriate generalizations, and the like.

Reasonableness of Conclusions/Recommendations. This criterion relates to judgments about those statements which go beyond simple interpretation of results. The consideration here is the degree to which a researcher is justified in "making something" of his findings. The evaluator should be alert to the "tightness" of these statements; that is, do they follow the general design? Are his conclusions substantiated? exaggerated? modest? Has he gone beyond his data? In general, the main issue is whether the discussion or the conclusions are related to the design, substantiated by the data, and generally logical.

Clarity of Presentation. For the most part, this criterion speaks for itself. It is also quite similar to the corresponding criterion for developmental products. The key consideration is the degree to which the effort has been logically organized and described in plain, straightforward language making it easy to follow and understand. The problems, concepts, hypotheses, conclusions, and so forth should be clearly and logically stated. In addition, the product should be so described as to make it completely comprehensible and, in appropriate types of research, replicable.

Potential Impact. This last criterion is essentially identical to the last developmental products criterion. In assessing potential impact, evaluators should ask to what extent the product has the potential for improving educational practice on a major scale. The basic question is to what extent the product is likely to effect a change in educational practice, or research, considering all the characteristics of the product and other factors which may influence the adoption and utilization of its concepts.

Chapter 4

INSTRUMENTATION

As a prelude to instrument development, a review of the rating scale literature was undertaken. According to Suchman (1950), all classification judgment is predicated on either itemized or non-itemized classification methodologies. Non-itemized classification is based upon scales which have simple nominal definitions. That is, a variable is simply named, and ratings on that variable are requested. Definition of the conceptual dimension is presumed to be self-evident in the label.

The problem with non-itemized classification is, of course, obvious. Differences in the semantic connotations, as well as denotations, of the variable label can result in a great deal of inter-rater variability. The semantic differential technique is one method that has been suggested to dimensionalize category labels.

Itemized classification is defined in terms of as many meaningful attributes as possible. As more and more specific items are added to the definition of the variable in question, the definition takes on a more and more precise meaning.

Judgment in itemized classification is based upon subordinate judgments made with regard to each of the definitional attributes. One approach at aggregating subordinate judgments is simply to summate the subordinate judgments. This is frequently the case in the use of checklists, composite scale scores, and the like.

The problems of classification based on subordinate item aggregation are twofold. First, the number of potential categorization items that exist for any single variable is unlimited. Thus, random item selection is by definition almost impossible to achieve, and there is no rationale for the differential inclusion of items. Secondly, assuming representative items have been selected, there are no rules for assigning weight to the item contributions to the aggregate score.

Practical applications of scaling principles seldom adhere to either of these two extreme theoretical positions, however. Application usually falls as a compromise somewhere between the two. If one abandons the notion of arithmetic combination of subordinate item scores, it is no longer necessary that the definitional items be faithfully representative of the total item universe. On the other hand if one is willing to select items in a reasonably representative, precise, and explicit way, one can gain considerably greater inter-rater reliability than he could if he persisted at the non-itemized extreme.

TYPES OF RATING SCALES

Assuming that the dimensions of evaluation have been specified, Guilford (1954) has indicated there are essentially five broad categories of rating scales. Two of these techniques are commonly associated with the itemized or aggregate judgment approach. They are the cumulated points and forced choice methods. The former was rejected as a methodology for the reasons previously cited. The forced choice method, or pair comparison method, is a procedure in which the items being evaluated can be rank ordered. With each panel evaluating no more than eight to ten products, it would have been a relatively easy task to use this methodology. This procedure would have been inappropriate, however, inasmuch as comparative assessment of only minimally similar products would have been theoretically meaningless.

The goal of the project was to establish procedures for the evaluation of products *vis a vis* an external standard, i.e., a hypothetical standard of "the mean of all products of a similar character." Of course, there is the implicit qualifier "within the experience of the evaluator."

The three other forms of rating scales identified by Guilford are numerical scales, graphic scales, and standard scales. Numerical scales, as the name implies, are scales wherein the individual's judgment is reflected as an ordinal position on a number scale. Graphic rating scales are, by analogy, scales where the individual's judgment is reflected by a position on a linear scale. A standard scale is a scale where the evaluator's judgment is reflected in the match of the item to be judged against one of a given set of standards.

Regarding numerical scales, Guilford suggests: 1) if the experimenter wants to achieve greater equality of psychological intervals between categories, he should attach verbal anchors to the numbers (the same is also true of graphic scales); 2) the use of negative rating numbers is not recommended; and 3) terminal categories should not be described too extremely.

Regarding graphic scales, vertical graphic scales are usually better than horizontal graphic scales because they allow cues to be long enough to be more meaningful, and cues can be localized at points along the line. For unsophisticated raters, the positive end of the scale should always be presented first. Descriptive phrases should be concentrated as much as possible at points on the line. To counteract the tendency to cluster ratings too near the middle of the scale, the steps between cues near the middle might be somewhat enlarged..

SCALE LENGTH

Regarding the number of points to use on a rating scale, Guilford suggests that consideration should be given to: 1) the use to which the evaluation results are to be ultimately put, and 2) the capacity of rating scale users to differentiate.

If the results of the evaluation are scheduled as input for complex mathematical or statistical treatment, as in research projects, then the primary limitation to be considered is the limitation of judges in making discriminations. With training, fairly extensive discriminations can be made. Guilford agrees with Champney and Marshall (1939) that the "optimal number of steps for the rater who is trained and interested may be as many as three times seven."

Non-statistical consideration of evaluation results is much more limited in the range of values it can accommodate. Miller (1957) has suggested that human beings have difficulty dealing with more than seven categories at any one point in time, and that for complex applications, the number is probably closer to five. Guilford (1954) has also argued that for untrained raters the maximum number of steps, for a single rating scale, is probably five.

In terms of the application of results to policy decision-making, differentiation into more than five groups (e.g., outstanding, well above average,

average, below average, and exceptionally poor) would probably be quite unnecessary. Indeed adequate policy decisions could probably be made on a three-point differentiation (e.g., well above average, average, well below average) if some leeway could be allowed at the boundaries of the three groups.

Finally, Guilford has noted that the average inter-rater reliability of rating scales is in the region of .55 to .60, and Symonds concluded as early as 1924 that seven steps were sufficient to optimize inter-rater reliability. "At this level of reliability more than seven categories increases inter-rater reliability by an amount that is so small that it does not pay for the extra effort involved."

SPECIAL CONSIDERATIONS

The prominent types of errors to be guarded against in scale utilization are: 1) errors of leniency, 2) errors of central tendency, 3) errors of reflected quality (the halo effect), 4) errors of logical relationship, 5) errors of proximity (ratings on scales that are physically adjacent tend to be correlated higher than more remote ones), and 6) errors of inadequate application (evaluators who have had training in the definitions of the criteria and instrument application produce more reliable ratings than untrained evaluators).

Regarding the use of rating scales as a method for evaluation, Guilford has written: "As compared with their nearest rivals, pair comparisons and the method of rank order, the rating scale methods have certain definite advantages and the results often compare very favorably with those from more accurate methods." Five advantages listed by Guilford are: 1) rating scales require less time, 2) the procedure is more interesting to the evaluators, 3) rating scale methods have a much wider range of application, 4) they can be used with raters who have had only minimal training, and 5) the results obtained are not significantly different from those obtained by more involved methodologies. Guilford concludes that "in view of the lack of better procedures, the rating method promises to find welcome use for many years to come" (1954, p. 297-298). Consequently it was decided to predicate the product evaluation system on a rating scale methodology.

In reviewing the rating scale literature, however, it became obvious that there was a major assumption implicit in most theoretical work on rating scale development. That was the assumption that the individual using the rating scale had the capacity to make relatively fine discriminations in judgment at a single point in time. This assumption is typically acceptable because of an implicit corollary assumption that the procedure involves the comparison of a well understood event to an internal norm. For example, in rating an individual's performance on a given task, it was assumed that the nature of the task is well known even though the individual and/or his typical performance might not be. The rating required is of performance on a well-defined and reasonably well-understood task, against the norm array of all other performances of all other individuals in the experience of the evaluator.

In the task at hand, however, the entity being evaluated is, by definition, a relatively new, and hopefully unique, entity which can be compared only to similar products in the experiential background of the evaluator. Thus it would be far less reasonable to expect an evaluator to make a highly differentiated response at a single point in time.

The situation seemed to call for a procedure analogous to the method of successive adjustments in psychophysics (Osgood, 1958). As far as could be determined, this method has no counterpart in psychometrics. In this procedure an evaluator would be called upon to first make an initial gross evaluation, and then, after tentative location of the product in a judgment zone, to make a finer adjustment. Thus, the task of R&D product evaluation would seem to call for a two-stage, successive judgments model.

THE SUCCESSIVE JUDGMENTS MODEL: A NEW APPROACH TO SCALE CONSTRUCTION

The successive judgments approach is a procedure often used by teachers and instructors when they are called upon to grade large numbers of term papers, essays, etc. The papers may be read quickly to identify whether the paper is "pretty good," "okay," or "not very good." The "pretty good"

papers are reread carefully to see whether they are still "just" pretty good, or "very" good. Similarly the poor papers are read next, to see whether they are just "middlin" poor or "awful."

While the method of successive adjustments is a widely used procedure in psychophysics, and, for most purposes, far superior to the single judgment method (method of single stimuli), a review of the major references on methodologies in psychometrics did not reveal a single reference to this two-stage methodology.

SCALE DEVELOPMENT

It was decided to develop product rating scales so as to combine as many of the positive attributes described by Guilford as possible. In particular, it was felt that each scale should have verbal anchors for each scale point and graphic as well as numerical properties.

One major consideration was whether the scale would presume equal intervals as on ordinary rating scales, or variable intervals as on standard score rating scales. The use of standard score judgments requires a certain psychometric sophistication on the part of the evaluator, especially if products tend toward the upper or lower extremes of the scale.

In view of the fact that many panel members may not have the technical background to fully appreciate the variable interval properties of standard score scales, it was decided to follow the more traditional rating procedure of equal intervals. Furthermore, the literature suggested that there would be no serious decrement in the reliability of ratings if this decision were followed.

When instrument development was started, copies of products for evaluation had not yet been received; thus, there was no way to ascertain just how "unique" they would be. In addition, inasmuch as the possible future operation of the system may involve the use of evaluation panels composed of individuals with only minimal background in measurement theory, it was decided to develop a two-stage as well as a more traditional single-stage instrument and try them both in the pilot test.

Finally, in addition to simple rating *per se*, it was held important to provide evaluators the opportunity for unsolicited written comments immediately following the rating on each criterion.

Suffice it to say that various forms of the single stage and double stage scales were developed and tried out until the physical format, the wording of anchors, etc. were sufficiently stable to warrant reasonable consistency of interpretation and application across users. This process spanned a period of approximately six months.

Figures 4 and 5 show examples of the single judgment and successive judgments formats respectively. Full copies of both types of instruments, as they were used in the pilot test, are presented in Appendix B.

Figure 4

EXAMPLE OF SINGLE JUDGMENT SCALE FORMAT

(Impact Criterion)

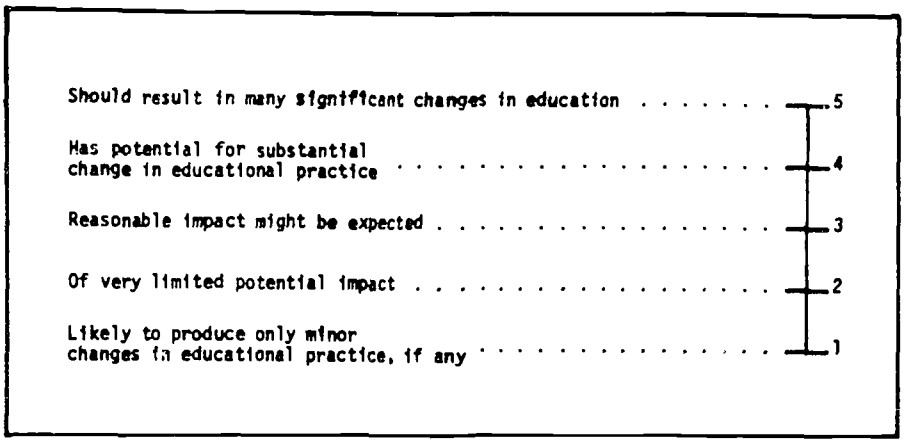
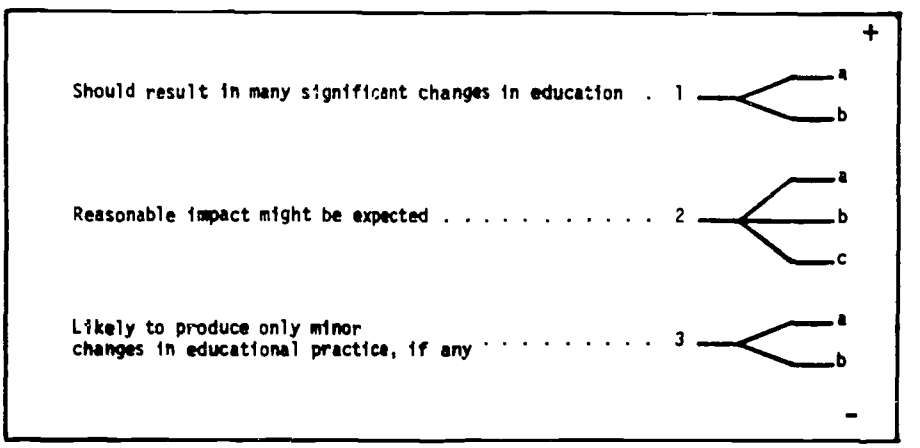


Figure 5

EXAMPLE OF SUCCESSIVE JUDGMENTS SCALE FORMAT

(Impact Criterion)



PART II

SYSTEM OPERATION

Chapter 5

PRODUCT IDENTIFICATION

To develop an evaluation system one must specify the domain of instances to which that system is to apply. If the goal is the evaluation of the product outcomes of laboratories and centers, one must specify the domain of those products.

Since laboratories and centers were funded to work on "problems of special significance to education" (see Bloom, 1968 and Chase, 1968), then it follows that their primary outputs should be solutions, or solution elements, for those problems. Product specification carries with it implicitly, then, the specification of the problem to which the product is purported to be a solution.

For purposes of this project, products were defined as proffered solutions to educational problems. This frame of reference was clearly the overriding one in the original foundation of R&D centers (Bloom, 1968) and was certainly the ultimate frame of reference used in the founding of the laboratory network (Chase, 1968).

Regarding product specification, it seemed most reasonable to have laboratories and centers themselves summarize the output they have generated in connection with the solution of the particular educational problems they have opted to work on. It was felt unreasonable to expect an external agent, regardless of how sophisticated, to properly infer the specific problems addressed by laboratories and centers. It was believed that the potential implications of problem identification and product evaluation were so crucial to an agency that they should not be delegated to a second or third party.

Accordingly, detailed instructions were given to laboratories and centers with regard to the particular frame of reference this project was using (namely, the definition of a product as a solution to an educational problem) and instructions and procedures were provided by which the appropriate scope of

the problem could be defined. Laboratory and center staff then identified the variety of product elements, i.e., outputs that had been generated toward the solution of that problem. This organized all of the production outputs into structured sets which together constituted the products of interest. It was these coordinated sets of elements which were then evaluated.

The development of the instructions and forms for this task involved consultation with selected laboratory and center directors and their key staff and underwent several cycles of empirical testing and revision during the spring of 1971. The final version, which was eventually adopted by NCERD as the foundation for their PARaDE reporting system, was discussed in detail with a representative sample of laboratory and center directors, approved by NCERD, and cleared for distribution on 22 October 1971. The instructions for product reporting, and the product reporting forms, are attached as Appendix A.

A total of 4,400 product reporting forms and 400 instruction booklets were eventually requested by, and distributed to, the 22 extant laboratories and centers.

FIELD TEST RESULTS

There was considerable variation in the degree to which the various agencies followed suggested guidelines with regard to product reporting. Some agencies opted to report their efforts in the most consolidated way and consequently reported relatively small numbers of fairly complex products. On the other hand, others opted to divide their complex products into subcomponents and report on each element separately. The number of products reported by individual laboratories ranged from 2-68 for developmental products and from 5-118 for knowledge products. The "size" of these products, however, ranged from materials costing less than a dollar (a 75¢ wall chart or a free brochure, for example) to complex, multi-media, individualized instructional systems costing many ten's of thousands of dollars.

The majority of the laboratories and centers responded promptly and conscientiously to the task. Several groups volunteered recommendations for the improvement of the procedure; two indicated they found the exercise beneficial for their long-range planning.

Several laboratories and centers found it difficult to meet the target submission dates, and extensions were arranged. In addition, four other laboratories indicated they felt they could not, or should not, comply with product reporting at all. Two of these were laboratories on terminal funding who, quite naturally, felt there would be little advantage, either to themselves or to the project, to complete reports. The other two felt they should not respond for a variety of local reasons.

Although over half the laboratories and centers expressed concern over the five-week time span allowed for completing the forms (the initial five week reporting period was eventually extended to ten), there were virtually no questions regarding how to fill out the forms.

The vast majority of product reports were also well within the space limits provided on the forms. Only occasionally was additional space required. Knowledge reports averaged approximately 2/3 of a single-spaced type-written page; developmental product reports averaged approximately a page and a half.

Figure 6 shows the distribution of reports by type of product, and developmental stage of the product. A total of 851 documents had been received as of January 3, 1972, the cut-off date for the field test. An additional 116 were received subsequently, raising the total to 967.

Figure 6

PRODUCT REPORT DOCUMENTS RECEIVED
AS OF MARCH 1, 1972

	Completed[1]	In Process	Subtotal 1/3/72	Received Subsequently	Total as of 3/1/72
Laboratories					
Knowledge	51	224	275	12	287
Developmental	52	127	179	36	215
University Centers					
Knowledge	110	213	323	56	379
Developmental	38	36	74	12	86
Totals, 1/3/72					
Knowledge	161	437	598	68	666
Developmental	90	163	253	48	301
TOTALS, 3/1/72	251	600	851	116	967

DISCUSSION

It should be pointed out, however, that number of documents is not synonymous with number of products. This is especially so in the case of knowledge products where, in certain cases, separate documents are used to report different elements of the same general knowledge product.

There are also other reasons why simple document counts cannot be used as product counts. For example, in some cases more than one report was filed by the same agency for the same product. In another instance, an agency reported three different editions of the same product as three different products. In still another instance, staff training materials, for internal use only, were reported as a developmental product.

[1] Knowledge products reported as completed but not published or otherwise made available to the professional public via some cataloging and reproduction service such as ERIC were considered as still in-process.

When copies of the products were requested for evaluation, in two instances their availability for evaluation was retracted. In another instance we were informed the product had been returned to in-process status because the product had been reconsidered and had been judged as needing further revision.

Several products appeared to have been completed prior to the initiation of the center reporting it; and in two other instances proprietary products were reported as products developed by the agency.

Almost twenty percent of the products selected for inclusion in the pilot test were sufficiently irregular to warrant some question as to their appropriateness for inclusion in the tryout.

In view of the variation in the judgments of respondents as to what items were appropriate for reporting, careful effort in any future implementation of the system (or in analysis of data currently in hand) should be directed to the validation of the data resulting from the product reporting procedure to insure equitable comparisons. The process of verifying the appropriateness of certain reports will, of course, be a matter of delicate interaction with agency directors.

In view of the great disparity across agencies in numbers of documents submitted, and in the range of types of instances on which documents were submitted, it is very clear that interpretation of raw data should be undertaken only very carefully. This should be especially the case in the interpretation of simple quantity data.

After strong admonition for caution regarding the danger of jumping to conclusions regarding the "number of documents submitted" and the inconsistent size of products reported, it is useful, nevertheless, to inspect the number of products reported.

Excluding five agencies which had not reported products as of January 3, 1972, and one agency which had reported only a single sample product, it can be seen from Figure 7 that 73 developmental products had been completed by the laboratory/center network since its implementation.

BREAKDOWN OF COMPLETED PRODUCTS/DOCUMENTS SUBMITTED AS OF 1/3/72

	KNOWLEDGE						DEVELOPMENTAL		KNOWLEDGE AND DEVELOPMENTAL	
	COMPLETED TYPE I		COMPLETED TYPE II		TOTAL		TOTAL		TOTAL	
	DOCS.	PRODS.	DOCS.	PRODS.	DOCS.	PRODS.	DOCS.	PRODS.	DOCS.	PRODS.
LABORATORIES										
1	1	1	2	2	3	3	2	2	5	5
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	28	21	27	23	55	44	14	11	69	55
6	2	1	7	4	9	5	3	3	12	8
7	0	0	0	0	0	0	0	0	0	0
8	8	6	6	4	14	10	15	14	29	24
9	8	4	60	31	68	35	14	6	82	41
10	3	3	7	7	10	10	2	2	12	12
11	0	0	0	0	0	0	0	0	0	0
12	1	1	1	1	2	2	2	2	4	4
UNIVERSITY CENTERS										
1	9	7	6	5	15	12	0	0	15	12
2	13	3	16	0	29	3	2	2	31	5
3	3	3	69	30	72	33	5	3	77	36
4	59	34	9	4	68	38	8	6	76	44
5	1	1	0	0	1	1	0	0	1	1
6	0	0	0	0	0	0	0	0	0	0
7	22	14	18	13	40	27	16	15	56	42
8	0	0	0	0	0	0	0	0	0	0
9	2	2	12	12	14	14	0	0	14	14
10	1	1	15	6	16	7	7	7	23	14
TOTAL LABS.	51	37	110	72	161	109	52	40	213	149
TOTAL UNIV. CTRS.	110	65	145	70	255	135	38	33	293	168
TOTAL	161	102	255	142	416	244	90	73	506	317

Figure 7

SUMMARY COUNTS OF COMPLETED AND IN-PROCESS DOCUMENTS RECEIVED AS OF 1/3/72,

AND DOCUMENTS RECEIVED AFTER 1/3/72

	KNOWLEDGE						DEVELOPMENTAL				TOTAL KNOWLEDGE & DEVELOPMENTAL			
	COMPLETED		IN PROCESS		LATE	TOTAL	COMPLETED	IN PROCESS	LATE	TOTAL	COMPLETED	IN PROCESS	LATE	TOTAL
	I	II	III	IV										
LABORATORIES														
1	1	2	2	2	0	7	2	10	0	12	5	14	0	19
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	28	27	4	25	0	84	14	54	0	68	69	83	0	152
6	2	7	2	2	0	13	3	5	0	8	12	9	0	21
7	0	0	0	0	12	12	0	0	18	18	0	0	30	30
8	8	6	0	0	0	14	15	0	15	30	29	0	15	44
9	8	60	21	29	0	118	14	28	0	42	82	78	0	160
10	3	7	22	2	0	34	2	30	3	35	12	54	3	69
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	1	2	1	0	5	2	0	0	2	4	3	0	7
UNIVERSITY CENTERS														
1	9	6	1	7	5	28	0	4	0	4	15	12	5	32
2	13	16	8	0	5	42	2	1	3	6	31	9	8	48
3	3	69	0	3	0	75	5	7	0	12	77	10	0	87
4	59	9	1	6	0	75	8	12	0	20	76	19	0	95
5	1	0	0	0	1	2	0	2	4	6	1	2	5	8
6	0	0	0	0	16	16	0	0	3	3	0	0	19	19
7	22	18	1	27	0	68	16	4	0	20	56	32	0	88
8	0	0	0	0	29	29	0	0	2	2	0	0	31	31
9	2	12	2	0	0	16	0	0	0	0	14	2	0	16
10	1	15	0	12	0	28	7	6	0	13	23	18	0	41
TOTAL LABS.	51	110	53	61	12	287	52	127	36	215	213	241	48	502
TOTAL UNIV. CTRS.	110	145	13	55	56	379	38	36	12	86	293	104	58	465
TOTAL	161	255	66	116	68	666	90	163	48	301	506	345	116	967

Figure 8

BREAKDOWN OF IN PROCESS DOCUMENTS AS OF 1/3/72

	KNOWLEDGE			DEVELOPMENTAL	KNOWLEDGE AND DEVELOPMENTAL
	IN PROCESS TYPE III	IN PROCESS TYPE IV	IN PROCESS TOTAL	IN PROCESS TOTAL	IN PROCESS TOTAL DOCUMENTS
LABORATORIES					
1	2	2	4	10	14
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	4	25	29	54	83
6	2	2	4	5	9
7	0	0	0	0	0
8	0	0	0	0	0
9	21	29	50	28	78
10	22	2	24	30	54
11	0	0	0	0	0
12	2	1	3	0	3
UNIVERSITY CENTERS					
1	1	7	8	4	12
2	8	0	8	1	9
3	0	3	3	7	10
4	1	6	7	12	19
5	0	0	0	2	2
6	0	0	0	0	0
7	1	27	28	4	32
8	0	0	0	0	0
9	2	0	2	0	2
10	0	12	12	6	18
TOTAL LABS	53	61	114	127	241
TOTAL UNIV. CTRS.	13	55	68	36	104
TOTAL	66	116	182	163	345

Figure 9

BREAKDOWN OF COMPLETED TYPE I KNOWLEDGE PRODUCTS AND COMPLETED DEVELOPMENTAL PRODUCTS BY MAJOR TAXONOMY CATEGORIES

	KNOWLEDGE PRODUCTS CATEGORY			DEVELOPMENTAL PRODUCTS CATEGORY			KNOWLEDGE & DEVELOPMENTAL CATEGORY			TOTAL
	I	II	III	I	II	III	I	II	III	
LABORATORIES										
	0	0	1	0	1	1	0	1	2	3
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	6	0	15	7	1	3	13	1	18	32
5	1	0	0	1	0	2	2	0	2	4
6	0	0	0	0	0	0	0	0	0	0
7	2	0	4	6	6	2	8	6	6	20
8	0	0	4	1	3	2	1	3	6	10
9	0	1	2	0	0	2	0	1	4	5
10	0	0	0	0	0	0	0	0	0	0
11	0	0	1	0	2	0	0	2	1	3
12										
UNIVERSITY CENTERS										
1	2	0	5	0	0	0	2	0	5	7
2	2	0	1	0	0	2	2	0	3	5
3	0	1	2	1	1	1	1	2	3	6
4	3	2	29	0	0	6	3	2	35	40
5	0	0	1	0	0	0	0	0	1	1
6	0	0	0	0	0	0	0	0	0	0
7	8	0	6	10	0	5	18	0	11	29
8	0	0	0	0	0	0	0	0	0	0
9	0	0	2	0	0	0	0	0	2	2
10	0	0	1	0	2	5	0	2	6	8
TOTAL LABS	9	1	27	15	13	12	24	14	39	77
TOTAL UNIV. CTRS.	15	3	47	11	3	19	26	6	66	98
TOTAL	24	4	74	26	16	31	50	20	105	175

Figure 10

A total of 102 knowledge products had been either published or otherwise made available through such channels as ERIC (Type I).[1]

Some 142 additional knowledge products had been produced and published in-house and were, presumably, retrievable by special request to the appropriate development agency (Type II). These latter documents, however, were not considered knowledge products for purposes of the evaluation system inasmuch as they did not meet the basic criteria for inclusion as a knowledge product, namely that a knowledge product must report (a) new knowledge, and (b) in a form that is readily available, i.e., retrievable, by other educational practitioners. Unlike technical papers filed with ERIC, where a permanent record copy is kept in archival storage, the contents of which are routinely abstracted, and reprints of which are made readily available, in-house publications and technical memoranda are not widely abstracted, if at all, and distribution is typically limited to only quantities in print. It is assumed that agencies would not have reported knowledge products as in-house publications if wider, more generally available, refereed publication of those products existed.

On inspection of Figure 7, if appropriate adjustments are made for the number of agencies reporting, it is interesting to note that there is no difference between laboratories and centers in the generation of Type II knowledge products, that is knowledge products published in-house. What is even more striking, however, is that there is no perceptible difference in the generation of developmental products as well. Of those laboratories and centers reporting, both types of institutions average approximately four to five completed developmental products for each institution. Thus it might be concluded that, assuming that there is no systematic bias in the strategies employed by laboratory and R&D center directors as to "quantity reporting," and there is no reason to believe there is, R&D centers compete very favorably with laboratories in the generation of developmental products.

---

[1] Based on a pro rata projection for the non-responding institutions, it is estimated that these figures reflect approximately 65% of the total laboratory/center network output. This estimate closely parallels the number of products reported by the CEDaR Information Office in its 1972 products catalog.

Further, R&D centers tend to produce nearly twice as many retrievable knowledge products, i.e., products published in some form accessible to the professional community.

If one looks at the relative distribution of interests of laboratories and centers, as shown in Figure 10, it is interesting to note that R&D centers produce approximately twice as many knowledge products in the area of the learner, the teacher, teacher-learner interaction, educational administration, and educational system development, than do laboratories.[1]

R&D centers also generate, on the average, more developmental products in the area of educational administration and educational systems development than laboratories, almost twice as many, and they generate comparable amounts of developmental products for dealing with the teacher, the pupil, and the teaching learning process.

In brief then, the surprising result of the analysis of "raw numbers" is that, on the average, R&D centers are not secondary to laboratories in the development of developmental products and they greatly exceed the laboratories in the number of published, and retrievable, knowledge products that they generate.

Two factors that have not been considered in this discussion, however, are the possible inequality of product unitization and the differential levels of agency support.

As suggested earlier, there is no reason to believe that research centers, as a group, systematically reported more atomistic products than laboratories. Both laboratories and centers reported products that were very large as well as products that were very small.

---

1 See Chapter 6 for the products taxonomy.

The second factor not discussed is the question of differential levels of agency support. For the past several years, for those agencies reporting products, laboratory support ranged from approximately 1 to 3.5 million dollars per year whereas research center support ranged from only .6 to .9 million dollars. The average 1971 funding for laboratories was more than 2 1/2 times that of centers. Mean aggregate funding, i.e., funding cumulative from the initial establishment of the agency, is on the same order. Since their inception, R&D centers have averaged a total of approximately 4.2 million dollars each, whereas laboratories have averaged a total of 8.1 million dollars each.

These data would seem to suggest, at least tentatively, that the critical mass notion of R&D funding is a fallacy, at least in the scale of expenditure of several millions of dollars per year. More modest funding extended over longer periods of time apparently accomplishes essentially the same net developmental result as mass funding over a shorter period of time, and with a higher probability of published research.

It must be borne in mind, however, that these findings refer to quantities where no consideration has yet been given to the relative quality of the products so produced. These conclusions are, of course, only of the most tenuous nature, and other possible factors have not been ruled out.

One point that is essential to repeat is that, if the proposed product evaluation system is to be implemented, an adequate, fair, and validated fix must be obtained on exactly what constitutes the real output of laboratories and centers. It will be essential to review and thoroughly assess the nature of the items reported as agency products.

Chapter 6

PRODUCT CLASSIFICATION

Although the pilot test of this project was to be concerned with only those products completed in the last two years, because of the relative sparsity of products, data for all products completed since the inception of the laboratories and centers were pooled.

A total of 73 developmental and 102 knowledge products were reported, as of January 3, 1972, as having been completed. These 102 knowledge products were comprised of 74 knowledge products reported as totally completed plus an additional 28 products not yet fully completed but for which some results (i.e., component studies) had been completed and reported.

"Completed" developmental products are those completed to the point where they were ready for transmission to the next agency in the developmental chain. "Completed" knowledge products are those published and retrievable through some standard topical indexing such as The Readers Guide to Periodical Literature, The Psychological Abstracts, Child Development Abstracts, Research in Education, or are accessible to the professional public via such "non-publication" channels as U.S. Government Reports, ERIC microfiche, Journal Supplement Abstract Service, etc.

THEORETICAL CONSIDERATIONS

Objects (products) are ordered (classified) for two reasons. One is to make the object array more comprehensible. The other is to permit the condensation of that array so that accommodations can be made to classes of objects rather than specific objects independently. How one classifies R&D products then is, in part, a function of one's functional perspective, i.e., how one defines product and what one wishes to do with them.

The identification of groups of highly similar products permits the selection of panels of appropriate evaluators to evaluate all of the products within

given groups, an option of important theoretical as well as economic advantage. It also permits an analysis and summarization of product evaluation data by product groups, or classes, once the individual products have been evaluated.

Thus, product classification should be engaged in prior to product evaluation for the purpose of evaluator selection. This prior product classification can then be used subsequent to product evaluation for the generation of summary evaluation statements.

MARKET-ORIENTED CLASSIFICATION MODELS

Three separate market-oriented models for product classification have been widely used. The first may be called a user or customer-oriented model, the second a production or accounting-oriented model, and the third a supplier or market-distribution model.

The User-Oriented Model. The problems with which teachers and principals are faced are coordination and management. Thus, they tend to be concerned with what the product is to do and how it is to be used. They are concerned with questions of target audience and the mechanics of implementation. Depending on which issue is paramount in their minds, they may consider products in terms of such categories as third grade spelling materials, fifth-grade social studies materials, cultural enrichment materials for inner-city children, etc. Or, conversely, they may categorize them as self-instructional materials, consumable materials, materials requiring teacher supervision, small group discussion materials, etc.

The Production Accounting Model. Product developers typically define products in terms of their discreteness as production items. Attention is focused on the component elements of the product. The level or method of application of the product seldom plays a role. Products generally are considered in terms of their physical characteristics, e.g., film-strips, textbooks, teacher guides, tape recordings, workbooks, audio-visual kits, etc. Each is an entity of production which eventually can have a unit price tag assigned to it. This type of classification is commonly seen in those large-scale

production efforts where close production monitoring must be maintained and where production cost accounting must be established.

The Supplier Model. From the point of view of the supplier/distributor (i.e., the point of view of sales and marketing), products should be, and are, typically defined in terms of the unit of supply, i.e., in terms of the items that have to be inventoried, priced, and distributed. The package to be supplied is usually a composite of a number of production items. Examples of this form of product definition are SRA Reading Kits, IPI Mathematics, and the Far West Laboratory Minicourses. Minicourse 1, for example, consists of eleven 16mm color-sound films, a teacher's handbook, a coordinator's handbook, a general information handbook, and a book of research readings. The "product" exists as a composite of these elements. All are necessary for the operation of the minicourse. They are supplied as a unit and priced accordingly ($1,475).

DEVELOPER-ORIENTED CLASSIFICATION MODELS

The common models just discussed were all carefully considered but were felt inadequate for project purposes. Three further alternative classification models were identified.

The Topological Model. From the point of view of someone charged with overall supervision or monitoring, products may also be defined from what might be called a topological or formal point of view. Here the question is on the general area of the outcome. It is often useful to know the relative distribution of effort going into different priority areas. Priorities may be defined either from a political or program policy perspective. Examples of priority areas may be target group areas, e.g., pre-school education, inner-city education, career education, or product emphasis areas, e.g., basic research, developmental research, hardware development, materials development, and the like.

An example of this approach can be found in Division I, "Primary outcomes of project activity" of the 1970 NCERD taxonomy. An even more intensive effort along this line may be found in Roger Levien's "Preliminary Plan for the NIE."

The Requisite Tasks Model. Continuing in the frame of reference of the management of R&D, products might also be defined in terms of their requisite tasks, and in terms of the network of functions necessary to accomplish those tasks. This is an especially useful approach when manpower needs are to be assessed and allocations made. All products requiring the same types of developer skills are treated equally regardless of the target audience for whom they are intended, the subject matter with which they deal, etc. An example of the requisite tasks approach to product identification is that of the Oregon Teaching Research Division's study of RDD&E activities.

That study identified 235 task activities generic to the production of educational research and development products and then analyzed a number of major R&D products accordingly.

One of the peculiarities of this point of view is that it focuses attention on the component tasks of the product and never actually on the product itself. It would be impossible, for example, to differentiate Sesame Street from Project HOPE or perhaps even from IPI. The superordinate (focal) product is simply taken as a given.

The Functions Analysis Model. Still another alternative approach to product classification is predicated on function analysis rather than task analysis. This approach is concerned primarily with questions of group dynamics and personal interaction. It is concerned with defining products in terms of the patterns of interpersonal process, social interaction, and management style associated with their production. This approach has typically been of interest to social psychologists and sociologists. (See Sieber and Lazarsfeld, 1966, for example.)

PROJECT NEEDS

None of the above was relevant to the project at hand, however. From the point of view of the educational policy maker, basic interest should be in what the product can do for society, that is, on the problems the product promises to solve, not in production monitoring, application, management, supply

and distribution, operation, the theoretical origins of products, or the like. The products taxonomy for product evaluation, then, should be, in effect, a problems taxonomy. And theoretically, the problems taxonomy ought to be the result of a systematic needs analysis.

From an *a priori* point of view there are, perhaps, only three major problems in education: a) our teaching is poor, b) our content is questionable, and c) we don't know how to improve our efforts.

Our teaching may be poor because we don't know enough about the teacher, the learner, or the teaching-learning process. Our content may be questionable because it is either wrong, irrelevant, or even disruptive, (i.e., it interferes with subsequent learning). We may be ineffectual in improving education because we don't know how to use well what we already have, create more efficient systems, or initiate and operate, i.e., administer, new systems once they have been created.

Assuming this, our needs are deceptively simple. We need more knowledge about basic processes and the optimum strategies for improving teaching, learning, curriculum selection, program administration, and the introduction and nurturance of innovation. We need better materials to use in our instructional efforts, i.e., better curricular and instructional support materials. We need better training in how to use the materials available. And we need assistance in the implementation of improved programs.

In other words, we need: a) more knowledge about teaching, learning, and curriculum administration; b) more tools, i.e., instructional materials, to use in teaching; c) training on how to use the new instructional materials; and d) assistance - often financial assistance - for the introduction of innovation.

THEORETICAL ISSUES IN TAXONOMY DEVELOPMENT

A useful classification system needs to be a) complete enough to assist in its expressed purpose, b) brief enough to be manageable, c) open enough to

admit new categories, d) explicit enough to allow reasonable reliability in classification, and e) sufficiently internally consistent (logical) to be valid (i.e., useful). There are many myths associated with the development of classification systems, however.

The principle of exhaustive classification is frequently held to be an essential constraint. The principle of exhaustive classification holds that all conceivable exemplars must be classifiable. While the goal of taxonomic universality is desirable, this principle is honored more in point of law than in spirit through the use of such residual categories as "other" or "not otherwise specified."

A second "essential" constraint is the principle of exclusive classification, the principle that an item may be classified in one and only one category. Taken together these two "principles" constrain classifications to completeness and mutual exclusiveness, i.e., universality and categorical independence.

The historical antecedents of these two principles derive from Aristotelian philosophy where absolutes and truths were fundamentals. The logic of contemporary science and mathematics is pragmatic, however, and exists in counterpoint to Aristotelianism.

The history of mathematics is a history of the accommodation of logical inconsistencies. To the extent possible inconsistencies were incorporated within the logic net of the existing arithmetic by the introduction of new, previously undefined, and previously unanticipated, concepts. The creation of imaginary numbers is a case in point of logical inconsistency being resolved by the invention of a new construct within the logic net.

The introduction of Boolean algebra and non-Euclidean geometries are examples of the creation of entirely new logic systems when prior systems could not be easily modified.

The best practical arguments for these two principles were user convenience, either convenience of data classification in the first instance, or confidence

of data retrieval in the second. This was essential in 19th and much early 20th Century science, but with computer technology it is just as easy today to have multiple classification systems as single classification systems. Witness, for example, the ERIC system, which classifies according to a thesaurus of descriptors and retrieves on the intersect of one or more classification descriptions. Nor are multiple classification systems necessarily contemporary. Bibliographical indexing, i.e., abstract topical indexing, has always used multiple classification as contrasted to the discrete classification methods characteristic of the early physical sciences.

Examples of the violation of mutual exclusivity are rife in all of the major taxonomic structures in science today. The two best known are the biological and the physical element taxonomies, although astronomical classification is currently in much greater and more rapid upheaval.

PRACTICAL PROBLEMS IN TAXONOMY DEVELOPMENT

There are, of course, practical problems, as well as theoretical problems to be considered in the development of any taxonomy. Taxonomies, to be useful, must be both reliable and valid, i.e., they must be sufficiently precise to permit similarity of classification over time, and they must be internally consistent enough to permit reasoned extrapolation.

Logical integrity in a classification system is valued because of its heuristic potential. Unfortunately, however, such integrity at times becomes an end in itself and, like over-zealousness for reliability alone, can compromise the system through reduction to logically rigorous, but extremely narrow, or even trivial and functionally useless, specification.

Beginning logic courses are rife with examples for students of logically derived propositions that are meaningless in application because of the narrowness of the logical system applied.

Logical rigor is most easily obtained through minimization of relational complexity. Relational complexity is minimized with the assumption of mutually

exclusive, independent categories. Concurrently, reliability is enhanced--hence the emphasis on the principles of exhaustive and exclusive classification mentioned earlier.

The point being made is that while it is desirable, all things equal, to have a taxonomic system which is exhaustive and mutually exclusive, the prime consideration for taxonomy adoption is its usefulness for the intended purpose, not its philosophical elegance.

TAXONOMY DEVELOPMENT

Upon receipt of the requested product information reports, samples were drawn and used to test the comprehensiveness and relevance of our *a priori* problems taxonomy. Because of the long history of classification of publications according to topical categories, it was felt that there would be less problem with the classification of knowledge products than of developmental products. Consequently, the early tests of the taxonomy were carried out with samples of developmental products.

The developmental products were ordered numerically and every fifth product was assigned to a sample group. The total domain of products was thus divided into five samples. The products in the first sample group were then classified according to the *a priori* taxonomy. Difficulties in the classification of products resulted in revision of the taxonomy and the process was repeated with the second group. The process was reiterated four times. By that time the taxonomy had stabilized. No changes were required for classification of the fifth sample. The process was repeated for knowledge products. Only three trials were required for stabilization of the taxonomy for knowledge products. Eight successive versions of the taxonomy were tested in this way.

The resultant taxonomy is a six-stage successively differentiating classification taxonomy. That is, it has a series of main headings which are differentiated into successively more and more specific categories.

There are 80 specific product categories subsumed under 16 general product classes. (There were originally 129 categories.) These 80 product classifications are the composite of 47 general product categories further differentiated into 42 sub- and sub-subcategories. The number of classification categories at each succeeding step in the taxonomy are: 3, 16, 47, 30, 6, 7 respectively. Or, if one combines all categories from the third level and below, excluding the redundancy of subordinate classification, the pattern is: 3, 16, 80. If one excludes all nonspecific categories such as "other" which are necessary to make the taxonomy exhaustive, the pattern is: 3, 13, 41, 26, 6, 6, or with combination, 3, 13, 69.

Not all products can be, or need be, classified to such a degree of specificity, though. The system is most complex for teacher training where all six levels of the taxonomy are used. It is next most complex in the areas of curriculum, instructional systems, assessment, and evaluation, where the taxonomy goes to four levels. It is least specific in the areas of the learner and learner characteristics, instructional methods, vocational education, pupil personnel services, general school administration, and procedures for product information dissemination and implementation. In those areas the taxonomy goes to only three levels of specificity.

The degree of specificity maintained in any area is, in part, a function of the precision with which the product was reported. Detailed specification was deleted from the final taxonomy where there were no products even remotely related to those categories. The maintenance of a highly complex taxonomic procedure for relatively few products only exacerbates problems of coder training and taxonomy use.

Although taxonomy revisions were primarily concerned with the elimination of low- or no-frequency categories and improving the conceptual specificity of those remaining, on occasion new categories were added. Even so, the 170 products reported as completed still required only 32 of the 69 non-residual categories.

Because of the open nature of the resultant taxonomy; i.e., because of its capacity for the admission of subordinate categories, detailed refinement can be reinstituted in the taxonomy as the need arises.

The taxonomy is summarized in Figure 11.

Part I of the taxonomy is used to classify new knowledge and/or developmental products about, or relevant to, the improvement of teaching and teacher training. This includes a better understanding of the personal-social characteristics of learners and of teachers, classroom management processes, the learning process, and the perceptual/cognitive motor processes underlying human learning, or on which human learning is based.

Part II of the taxonomy is used to classify knowledge and developmental products concerned with the curriculum, its structure, organization, requisite sequencing, methodology, and materials, including workbooks, teacher's guides, filmstrips, audio-visual aids, programmed materials, and the like. Part 2 excludes hardware and hardware operating or utilization manuals.

Part III is used to classify products concerned with the creation, improvement, evaluation, and/or management of educational research and development, instructional systems, public school programs, college programs, school business operations, the dissemination and implementation of new products and practices, and the like.

PRODUCT CLASSIFICATION

After finalization of the taxonomy, all products were then classified on the basis of information provided on the Product Reporting Form. Each developmental product was coded by four independent coders. Each knowledge product was coded by three independent coders.

Coding Rate. An experienced product coder, i.e., a coder who has had a minimum of two half-days training and supervised coding experience, requires approximately two minutes to code a knowledge product and three minutes to

Figure 11

PRODUCT CLASSIFICATION TAXONOMY

Product Counts*			Taxonomy Categories
Know-ledge	Develop-mental	Total	
24	26	50	I. LEARNING - TEACHING
7	0	7	A. CHARACTERISTICS OF THE LEARNER AND OF THE LEARNING PROCESS
-	-	-	1. Personal competencies
2	0	2	2. Socio-emotional foundations
2	0	2	3. Perceptual/cognitive foundations; achievement
3	0	3	4. All, some, or other in the above
17	26	43	B. INSTRUCTIONAL MANAGEMENT: TEACHERS, TEACHER-PUPIL INTERACTION, AND TEACHER TRAINING
7	12	19	1. General teaching skills
-	-	-	a. Planning
6	8	14	b. Operation
-	-	-	c. Learner progress assessment
1	4	5	d. All, some, or other
6	4	10	2. Teacher characteristics/personal skills
0	10	10	3. Specific techniques
0	3	3	a. Use of specific instructional materials in:
0	1	1	i. Basic abilities
0	2	2	ii. Academic programs
-	-	-	a. Math
0	1	1	b. Science
-	1	1	c. Reading
-	-	-	d. Literature/writing/composition public speaking
-	-	-	e. Social studies
-	-	-	f. Foreign language
-	-	-	g. Other
-	-	-	iii. Cultural/leisure
-	-	-	iv. Civic/citizenship
-	-	-	b. Use of hardware/computers/special equipment
0	1	1	c. Use of new school/classroom organizational patterns
0	6	6	d. Improvement of teaching in specific content areas
0	3	3	i. Via improved teacher content knowledge
0	3	3	ii. Via improved general teaching skills
4	0	4	4. All, some, or other

*Received as of 1/3/72.

Figure 11 (continued)

PRODUCT CLASSIFICATION TAXONOMY

Product Counts			Taxonomy Categories
Know-ledge	Develop-mental	Total	
4	15*	19	II. CURRICULUM - CURRICULUM MATERIALS
1	0	1	A. STRUCTURE AND METHODS
-	-	-	1. Learning hierarchies
-	-	-	2. Topical hierarchies
-	-	-	3. Methods
1	0	1	4. All, some, or other of the above
0	4	4	B. BASIC ABILITIES
0	2	2	1. Self-management skills
0	1	1	2. Social skills/affective development
-	-	-	3. Process skills
-	1	1	4. All, some, or other
2	7	9	C. ACADEMIC PROGRAMS
2	7	9	1. Content learning
1	1	2	a. Math
0	2	2	b. Science
-	-	-	c. Reading
0	2	2	d. Literature/writing/composition/ public speaking
0	1	1	e. Social studies
1	1	2	f. Foreign language; english as second language
-	-	-	g. Other
-	-	-	2. Cultural/leisure/general enrichment programs
-	-	-	a. Cultural/avocational/hobby/aesthetic
-	-	-	b. Athletic
-	-	-	c. Citizenship/civic/public service
1	4	5	D. MANUAL ARTS/BUSINESS/HOME ARTS/AND VOCATIONAL TRAINING PROGRAMS
0	2	2	1. Info re. world of work/vocational - career information programs -
0	2	2	2. School training programs, e.g., wood shop; home economics
-	-	-	3. OJT or preemployment specific job training programs - for actual employment training
1	-	1	4. Associated job relevant skills (prevocational)
-	-	-	a. Locating jobs and job opportunities
-	-	-	b. Retaining newly acquired jobs
-	-	-	c. Changing jobs
-	-	-	E. OTHER

*One developmental product not included in this summary due to multiple classification.

Figure 11 (continued)

PRODUCT CLASSIFICATION TAXONOMY

Product Counts: Knowledge	Product Counts: Developmental	Product Counts: Total	Taxonomy Categories
72*	31	103	III. SCHOOL ADMINISTRATION AND EDUCATIONAL R&D
6	0	6	A. GENERAL MANAGEMENT AND CONDUCT OF ED. R&D
5	0	5	1. General management: procedures, strategies
1	0	1	2. Requirements for development of specific products
32	16	48	B. INSTRUCTIONAL/EDUCATIONAL SYSTEMS
1	0	1	1. Instructional management: general
0	1	1	2. Information systems/student record files
29	12	41	3. Goals/analysis/assessment/accountability
1	1	2	a. Objectives
12	8	20	b. Tests/test development/instrument development
14	1	15	c. Evaluation
2	0	2	d. Statistics, measurement theory
-	-	-	e. Systems: theory, operations research
0	2	2	f. All, some, other
2	3	5	4. Specialized components
2	3	5	a. Equipment/hardware/software development and/or utilization
-	-	-	b. Procedures for improving supply/logistics
-	-	-	c. Other
1	1	2	C. PUPIL PERSONNEL SERVICES
1	0	1	1. Guidance/counseling
-	-	-	2. Psychiatric/psychological diagnosis
-	-	-	3. Therapy
0	1	1	4. Other
4	0	4	D. BUSINESS OPERATIONS
-	-	-	1. Personnel management
3	0	3	2. Financial management
-	-	-	3. Physical plant management
1	0	1	4. Public relations/cooperation
-	-	-	5. Production
23	6	29	E. GENERAL MANAGEMENT: OTHER
1	4	5	1. Operational programs
6	0	6	2. Public schools
3	2	5	3. College administration
-	-	-	4. Manager training/inter- intra personal skills
7	0	7	5. Group dynamics, influence patterns, organizational behavior
6	0	6	6. Other

*Two knowledge products not included in this summary due to multiple classification.

Figure 11 (continued)

PRODUCT CLASSIFICATION TAXONOMY

Product Counts			Taxonomy Categories
Know-ledge	Develop-mental	Total	
4	8	12	F. DISSEMINATION
2	2	4	1. General information dissemination: theory, procedures
2	6	8	2. Specific product information dissemination
1	0	1	G. IMPLEMENTATION
1	0	1	1. Adoption of new techniques/procedures (change agent functions)
-	-	-	2. Maintenance and exportation of innovations
1	0	1	H. OTHER

code a developmental product. The time difference is primarily a function of the amount of information that must be read from the product form. A sustained coding rate of 30 knowledge products or 20 developmental products per hour was characteristic for hour-long coding sessions. With modest experience in use of the taxonomy, a sustained coding rate in excess of this can probably be attained.

Assuming that all product information was complete, that knowledge product reports had been combined appropriately, and all product irregularities had been resolved, it is reasonable to expect that the entire array of in-process as well as completed products could be coded in the equivalent of approximately two man-weeks. The time lapse would be somewhat in excess of two weeks, however, inasmuch as the physical fatigue and tedium factor is such that a single individual should not be asked to code products for more than perhaps two hours a day.

Reliability. In general, the literature on taxonomy development and utilization is marked by an almost total lack of empirical attention to the question of the reliability of the taxonomy, or, to put it more accurately, to the degree to which exemplars can be reliably coded according to the categories of the taxonomy.

This would seem to be an important question, as the practical utility of a taxonomy would be severely restricted if it could not be used effectively to classify products.

If exemplars cannot be reliably classified according to the categories, then the taxonomy technically ceases to exist as a classificatory device and degenerates to a simple partitioning device. (Reliability is much less critical in such high technology systems as ERIC inasmuch as products are multiply classified. In such cases, descriptor intersects are highly overdetermined.)

Five independent tests of the reliability of codifying products were made; three for developmental products and two for knowledge products. Inter-rater

reliability was defined in terms of the percent of products for which there was complete agreement regarding taxonomic classification, across all coders. These results are summarized in Figure 12. In brief the reliability of classifying developmental products into the detailed taxonomy categories was on the order of .85 for developmental products and .65 for knowledge products.

From a more practical point of view, i.e., from the point of view of classifying products for assignment to evaluation panels, the purpose of product classification in the first place, reliabilities averaged .92 for developmental products and .88 for knowledge products.

Overall, across five different independent reliability checks, coding a total of 93 knowledge and developmental products, there was 90.4 percent agreement (i.e., total consensus) across all independent coders as to the proper assignment of products to evaluation panels. There was 77 percent unanimity among product coders as to the precise, detailed topical designation of the product. The latter are understandably lower than panel assignment reliabilities because of the much greater detail required for complete taxonomic classification. The taxonomy in places goes to six levels, a degree of specificity not needed for the designation of evaluation panels or the assignment of products to panels. Further, some products do not have a single predominant topic. Some teacher training products, for example, deal with teacher characteristics, interpersonal communication skills, and specific techniques for classroom management, without giving any indication which is the primary focus of the materials. Consequently, on such occasions, conflicting coding in the fourth or fifth levels of specificity can easily occur even though there would be intercoder consensus as to a more general classification and as to which evaluation panel it should be assigned.

Figure 12

RELIABILITY OF PRODUCT CLASSIFICATION
ACCORDING TO TAXONOMY CATEGORIES

Trial*	E	F	G	J	L	Composite
Consensus Re. Panel	100	88	89	95	80	90.4%
Taxonomy Consensus	89	88	78	75	55	77.0%
Number of Coders	4	4	4	3	3	
Number of Products	18	17	18	20	20	
Type of Product	D	D	D	K	K	

*Trials A, B, C, D were used to revise taxonomy for developmental products.
Trials E, F, G were approximately 30% stratified sample of developmental products.
Trials H, I, K were used to revise the taxonomy for knowledge products.
Trials J, L were approximately 20% stratified sample of knowledge products.

Results. A total of 175 products were reported as completed. Of these, 172 were coded on a single, clearly predominant, taxonomic category; 3 received multiple classification.

Figure 13 summarizes the number of products by the three major sections of the taxonomy.

Figure 13

COMPLETED PRODUCTS
BY MAJOR TAXONOMY SECTIONS

Type of Product	Category I Teaching/Learning	Category II Curriculum	Category III Educational R&D and Administration	Totals
Knowledge	24	4	74	102
Developmental	26	16	31	73
Totals	50	20	105	175

Figure 11, presented earlier, indicates the number of completed knowledge and developmental products per taxonomic category. The preponderance of products (57%) cluster around only three areas: 1) teacher training; 2) objectives, tests, and test development, and 3) school/college administration.

It should be recalled, however, that due to non-reporting on the part of some agencies, the product domain for this project constituted only an estimated 65% of the total laboratory and center output. Whether these same relative distributions would be maintained across the total product domain is a matter of question.

Inasmuch as the sums for each set of subordinate categories are repeated in the totals for superordinate categories, care must be exercised in combining totals across categories representing different levels of specificity.

Chapter 7

EVALUATOR SELECTION AND TRAINING

After the domain of completed products has been identified, and products have been classified into homogeneous taxonomic groups numbering, to the extent possible, some eight to ten products per group, independent panels of judges are then formed to evaluate all of the products within each group.

It is self-evident that products should be judged by the most knowledgeable evaluators possible. The panel of evaluators should be composed of individuals knowledgeable about user needs and concerns, subject matter specialists, product developers, and evaluation specialists.

If products are intended for a special ethnic group, then the evaluation panel should also have representation from that group. For example, if materials to be evaluated deal with bilingual programs, then the ethnic group for whom the bilingual programs are intended should be represented on the evaluation panel. This requirement is simply an extension of the criterion of user representation.

User representation is interpreted as representation on the part of school personnel, not child representation. Should the situation seem to warrant it, however, actual learner representation might also be appropriate and should also be considered a possible option by the evaluation coordinator. The judgments of such adjunct panel members should serve as inputs to the final deliberations of the core panel members.

The evaluator nomination, selection and training procedures are described in the following paragraphs.

EVALUATOR SELECTION

Laboratory and center directors are requested to nominate panel members for each product area in which they will have products evaluated. If a laboratory or

center has produced a product for a special ethnic group or other special target group, they should nominate special group representatives at the time they nominate subject, evaluation, and user experts.

A request for evaluator nominations for all required panels is also made of the governmental staff responsible for the administration of the laboratory and center program, and from the past-presidents, presidents, vice presidents, presidents-elect, and executive committees of AERA, and APA Divisions 15 and 16, and other national professional organizations as deemed appropriate by the evaluation coordinator.

While this procedure, on the face of it, would appear to be quite involved, it should be remembered that this nomination process need be conducted only once every two or three years, and once the current backlog of completed products is evaluated, will involve only relatively few panels at any one time.

The rationale for using such a large nomination base is that, through cross tabulation and winnowing, only those who receive nominations from a variety of sources would be retained. Such individuals, presumably, would be relatively prominent in their disciplines.

This procedure was tested using three product groups. The product areas were: 1) educational/instructional systems, 2) vocational/career education, and 3) child development/human learning/early childhood education.

It was originally expected that laboratories and centers would be quite eager to nominate potential evaluators for their products, and that there would be lesser interest in providing nominations on the part of the elected officers of professional organizations. This expectation did not appear to be justified, however.

There was no requirement that nominators identify themselves or their agencies; thus there was no possibility to systematically identify those who did and who did not nominate. It was noted, however, that of those who voluntarily did so, an unduly large proportion was the elected officials of professional organizations and unaffiliated with laboratories or centers.

The most surprising finding, however, was the almost total idiosyncrasy of the nominations. Figure 14 summarizes the number of individuals nominated by each of the three subject areas and the frequency of the individual's nomination.

Figure 14

EVALUATOR NOMINATION RESULTS

	Educational Systems Panel I	Vocational Education/Training Panel II	Early Childhood Human Learning Child Development Panel III
Number of Nominations	60	58	68
Number of Unique Nominations	56	56	59
Number of Individuals Nominated 2 times	4	2	7
Number of Individuals Nominated 3 times	0	0	2
Number of Individuals Nominated 4 times or more	0	0	0

Of the total of 186 individuals nominated for the three panels, 92% or 171 individuals, were nominated once and only once. Only two people received more than two nominations. No one was nominated more than three times.

These results were startling, to say the least, for they suggest, on the face of it, considerable confusion within the field as to who actually constitutes the professional leadership. (It is assumed, of course, that there is a recognizable body of experts in the subject area.) Why this should be the case is hard to explain.

An analysis of most frequently cited authors in technical and scholarly publications yields a fairly small, highly visible coterie of experts. It may be that nominators are hesitant to assume that the extremely prominent leaders in the field would be willing to serve. It may be that the nominators, some of whom are already acknowledged leaders in the field, hesitate to nominate themselves. It may mean that the field is so broad that there are more leaders than anyone imagined. Or it may mean that many nominators are just simply uninformed about leadership in the technical areas.

It is also interesting to note the striking lack of nomination of individuals who have served previously as laboratory and center site-visit evaluators. With but few exceptions, individuals used as laboratory and center evaluators in the past were not nominated as proposed product evaluators. Whether this is an artifact of the narrowly defined subject matter content of the three product areas selected (only 15% of the total completed product areas were involved) or whether it is a condition that will continue to obtain when nominations for the remainder of the product areas are requested will remain to be seen.

Upon careful inspection of these lists, however, one cannot help but be struck by the number of names of professionally prominent individuals associated with those subject matter areas who were not nominated. This, plus the strikingly low inter-nominator consensus gave rise to the insertion of a new step, not originally anticipated, into the evaluator selection procedure.

As a backup procedure, the evaluation coordinator should also make nominations for the evaluation panels. The coordinator should survey the major technical publications in the appropriate areas and generate lists of the editors, consulting and/or advising editors, and most frequent authors. He should also list, when appropriate, the names of current and recent elected officers of appropriate national organizations such as American Council of Teachers of Mathematics, American Personnel and Guidance Association, etc. If this does not yield a sufficient number of alternatives, as a final resort the evaluation coordinator should then turn to the senior membership lists of appropriate professional organizations and make nominations from among the senior fellow lists.

Alphabetical lists of the nominees and their institutional affiliations should be generated for each product area and circulated back to the laboratory and center directors for their review and critique.

Agency directors should be requested to indicate those evaluators they especially endorse and those evaluators about whom they would have serious reservation should they evaluate one of their products. These agency director returns are then codified and resultant lists generated in the following way.

Evaluator lists are generated such that those individuals most frequently nominated for a product area head the list for that area. Precedence within frequency categories is given to those individuals strongly endorsed by agency directors. Individuals for whom some laboratories have reported serious misgivings regarding their suitability as a product evaluator for products they have developed, are disqualified from evaluating products generated by that laboratory.

After the list is generated, vitae are obtained from American Men of Science, Leaders in Education, Professional Association Membership Directories, or the like.

In the event an evaluator has been nominated who cannot be located in any of the current editions of standard biographical references, he is dropped from the list. This is not to imply that the individual would be an inappropriate

product evaluator; only that, inasmuch as a variety of groups may be interested in the professional background of these product evaluators, their professional credentials should be a matter of public record retrievable from standard biographical sources.

The first 30 names on this list comprise the pool of potential product evaluators. The remaining names are kept in reserve for possible future use.

The initial 30 names, along with their biographical qualifications, are then submitted to OE for review. If any names on the proposed list are unacceptable to OE, they are deleted. The list of the evaluators remaining, after OE review, constitute the basic list from which panel members are drawn.

Panels are then selected so that at least 50% of the panel is composed of subject matter experts plus a minimum of at least one evaluator, one user representative, and, if needed, one target group representative.

EVALUATOR TRAINING

After a set of six to nine individuals agrees to serve in the evaluation of all of the products in a specified set, arrangements are made for an evaluator conference and training session. To the extent possible, at least two panels at a time should be convened for this session.

Fairly large numbers of evaluators could, of course, be convened for the conference and training session. At one extreme all evaluators could be convened at the same time. There are several arguments against attempting to maximize the number of trainees at the conference, however. For one thing, it would be extremely difficult to find common times when all evaluators could be there. The higher the number of individuals who are to attend the conference, the higher the absenteeism can be expected to be.

Secondly, the higher the number of conference participants, the less personal interaction can be expected to take place, especially *vis a vis* training on instrument utilization and on protocol and procedure discussions.

Thirdly, the larger the number of conferees, the less relevant the training examples will be to their areas of specialty.

Inasmuch as conferee time and travel will constitute fixed costs regardless of the number of conferences conducted, the only additional cost resulting from the conduct of a series of individualized training conferences, as contrasted to a single large training conference, is the cost of the evaluation coordinator's staff time. It is suggested this would be relatively modest in comparison to the disadvantages of relatively large group conferences.

A conference should be scheduled to take an entire day. There is always a tendency on the part of some to arrive late and leave early. It should be made quite clear that the training conference starts promptly at the designated time in the morning and that conferees should plan to arrive the night before. Reinforcement of this can be made by scheduling the distribution of materials, such as the evaluator's manual (see Appendix B), the agenda for the conference, product reporting sheets, product reporting instruction booklets, rating scales, etc. the night before the conference so that the evaluators can review them, if they wish, prior to the conference. It may also be helpful to schedule an informal social gathering immediately following the evening assembly.

Attention to such detail as conference luncheon plans will minimize straggling and will help to keep the conference on schedule.

At the beginning of the morning session, attention should be focused on the presentation of details of the evaluation system, the rationale underlying the evaluation procedure, general management procedures, future operations, and the like. Attention should then be directed to a detailed and thorough discussion of the criteria, and finally to the instruments to be used.

Even though materials will have been distributed the evening before and many, perhaps the majority, of the conferees will have read them, it is still important to work through these materials explicitly and thoroughly in the training session.

Upon completion of the discussion of the rationale, criteria, methods, and procedures, the evaluators should be walked through a sample evaluation of a

simple product. This should be a step-by-step consideration of a sample product, with ancillary discussion of the product and the criteria applications, as needed.

The sample product should be selected to be: 1) relevant to the professional areas of the evaluators being trained, 2) simple enough to be comprehended in a brief setting, and 3) heuristic enough to lend itself to discussion of relevant issues regarding the criteria, the judgment process, and the data recording procedures.

If possible, this first product should be evaluated before lunch, but after completion of the formal systems discussion. It is most desirable that this initial product evaluation be seen as an extension of the discussion of criteria definitions.

After lunch, two to four additional sample products should be evaluated. This should include one, or preferably two, relatively complex products. These complex products should be discussed and evaluated, at least hypothetically, for purposes of training, even though a thorough review of a complex product would be impossible.

At the conclusion of the training session, each evaluator should receive his projected evaluation schedule, a spare copy of the evaluator's manual, a supply of product evaluation forms, a supply of coordinator-addressed return envelopes for the return of the product evaluation forms to the evaluation coordinator, and copies of all of the simultaneous review products he is to evaluate.

Although the training conference can be held at any mutually convenient site, if there is to be one or more common site product evaluations, it would probably be most convenient to schedule the training conference and the group evaluation at the evaluation coordinator's home office so that the product evaluations can take place immediately following the training conference. This will minimize travel expenses and scheduling problems for a subsequent group meeting and will permit the evaluation of what will be the potentially more complex, and expensive, products while the training is still fresh.

If in the following year it becomes necessary to evaluate another set of products in the same general content area, and the same panel agrees to serve, it would not be necessary to replicate the training conference. If it were necessary to replace a few members, the replacements should join the training conference of one of the other product groups.

Chapter 8

PRODUCT PROCUREMENT

In addition to his other tasks, the evaluation coordinator has three main responsibilities directly pertaining to evaluation *per se*: to procure the necessary products, to coordinate evaluator activities, and to report results of the evaluations. The guidelines proposed in this and the subsequent chapters are based on the experience gained from the system tryout. This chapter describes the tasks subsumed under the first of these three responsibilities. The following chapter describes the tasks subsumed under the latter two responsibilities.

INITIAL INQUIRY

Arrangements for procuring products should be initiated well in advance of evaluator training in order to allow the coordinator sufficient time to review the products, to determine the mode of evaluation for each, and to make necessary equipment and scheduling arrangements. These activities are discussed in more detail in the following paragraphs.

Once the products to be evaluated have been identified, the evaluation coordinator must determine exactly what the products consist of, so that he can structure the evaluation procedure accordingly. The Product Report Forms provide some clues in this area. The descriptions indicate the different elements of a product, but in some cases the product descriptions may not be sufficient to indicate the level of effort that will be required to review them.

The evaluation coordinator should contact, in writing, the director of each agency whose products will be reviewed and request a single copy of each product. In the case of products too complex to mail, complete descriptive information about the product should be requested. It is important that the kinds of information desired be carefully delineated. Requests simply for "descriptive information" typically net only PR brochures which typically say less than the report form.

At the time the single copies are requested, information should also be obtained regarding the quantity availability of the product. Can ten copies be obtained (one for the evaluation coordinator and the remainder for the evaluators)? If not, where can the product be observed; or can a special demonstration be arranged?

The information on the Product Report Form should also be validated during this same contact. Several instances of inaccurate reporting were uncovered during the tryout of the evaluation system. It is important that the product reporting information be validated as early as possible. Based on the field test experience it is probable that the information provided on some 18-20% of the product reports may be questionable.

Upon receipt of the products (or detailed descriptive information), the coordinator should carefully examine the product and decide which mode of evaluation would be most appropriate. In addition, he should note factors which might need clarification for the evaluators. For example, in the evaluation system tryout it was often difficult to determine just how the pieces of more complex products fitted together.

One product included a slide-tape orientation to the product. Whenever possible, agencies should be encouraged to provide similar guides to their products if they think they would be of use.

A second area in which confusion may occur is between the product itself and the statements made about it on the product reports, particularly the problem statement. The pilot test evaluators often felt the problem, as stated on the form, was different from the one actually addressed by the product, and that many statements made on the form could not be supported. Thus, the evaluation coordinator should carefully review the Product Report Forms to identify any areas requiring further clarification or supporting documentation from the developer.

Several items of specific information should be included in the coordinator's letter to the agency director requesting the products. First, it should explain why the products are being requested and what will be done with them. The explanation should cover why these specific products are being requested.

Second, the letter should delineate in detail exactly what is being requested. Simply asking for a complex product such as "the XYZ Program" or even a not so complex product as "the ABC Kit" may not be sufficient. Products are often comprised of several elements and there is a pronounced tendency to provide only those elements which are most convenient for distribution.

Similarly, if documents supporting statements made on the Product Report Form or some form of descriptive guide to the product are desired, they also should be requested specifically. This is especially true of field test evaluation documents on which agency claims for product effectiveness are made. A general invitation to submit support documents resulted, in the tryout, in documents being supplied for only three of the twenty products reviewed.

In some cases a product may still be in the process of being published and, thus, not available in its final form. When this occurs, the evaluation coordinator should request copies of the prototype submitted for publication.

Third, information about shipping the products should be provided. The date by which they should be received should be indicated, as well as suggestions for the method of sending it, such as whether airport pick-up and delivery make air freight feasible or, if the mails are used, sending the material first class, registered, etc.

During the system tryout, several agencies did not send the materials by the date indicated. It is advisable, therefore, to allow a week or two of lead-time between the date indicated and the date on which they will actually be needed, so that delays in transit, which are likely, can be absorbed without jeopardizing the evaluation.

Fourth, a copy of the most recently submitted Product Report Form on the requested product should be included with the product request. The agency director should be requested to review the form and either confirm the accuracy of information on the form or update it. If any responses on the form seem unclear or self-contradictory, they should be noted and director clarification requested.

Fifth, where support equipment, such as tape recorders or videotape equipment will be required to review the product, the agency should be asked to provide complete specifications regarding the type and, if necessary, the model of equipment that will be needed.

Sixth, agencies should be asked whether or not they wish the products returned.

Finally, for products requiring some form of panel visit, information regarding the various locations where the product might be seen should be verified. If the agency indicates that a product is available from a specific marketing agency or at a specific location, the availability of the product at that location should be carefully verified before visits or conferences are planned.

As a result of this initial inquiry, the evaluation coordinator should know the composition of each of the products, how many copies can be obtained, what special arrangements, if any, should be made for reviewing the product, and what additional supporting information can be provided by the agencies.

ORDERING PRODUCTS

As the evaluation coordinator identifies what quantities of the products will be needed, requests to obtain the products can be initiated. Even under optimum conditions, as many as four to six weeks may be required for obtaining products. Up to two weeks may be needed by the responding agency just to prepare the material for shipment if a product must be assembled. Another

two weeks are often required to ship the materials, particularly when the products are too bulky to send via the mails. The requests should thus be made at least two, and preferably three, months in advance of the scheduled evaluation to allow for receipt and processing of the products. Shipment by means other than the U.S. mails, e.g., Greyhound Bus, United Parcel, Air Freight, or various airlines' "Next Flight Out" package services should be considered.

Three weeks after the letter of request has been sent out (assuming the agencies were given four weeks to submit the products), a follow-up letter should be sent to those agencies from which products have not been received. A second follow-up, by telephone, should be instituted when the "deadline" arrives if products are still outstanding. These follow-ups will serve as reminders to the agencies as well as provide information on the status of the product.

LOGISTICS CONTROL

In order to keep track of the various products during the evaluation, some form of product monitoring must be established. This can be as simple as a status chart maintained on a bulletin board or it can be a more complex procedure such as an IBM 407 accounting machine inventory control procedure or a McBee edge-punched card sort system. The form is not important unless large numbers of products, 40 or 50, or more, must be monitored in a very narrow time frame, e.g., 6-8 weeks.

Beginning with the initial inquiry, records should be made of the status of each product, including the following pieces of information:

- product title;
- developing agency;
- where the product can be obtained (if different);
- what the product consists of;
- what product elements and support materials were requested and when;

- what product elements and support materials were received and when;
- who was contacted to request the materials;
- who sent the materials (if different);
- special conditions, such as inaccurate reporting on the report form; and
- what mode of evaluation will be used to review the product.

These files should be augmented, as the actual evaluation begins, to include:

- who has reviewed the product;
- where it is currently located;
- when it has been returned to the evaluation coordinator; and
- when it has been returned to the developing agency.

In this way it should be easy to tell at a glance what the status of a given product is.

Because a relatively small number of products was dealt with during the tryout of the system, a simple, manually-posted log book was maintained. However, when more than 20 or 30 products are being evaluated, a simple log book system would be cumbersome.

As products are received, the materials should be carefully inspected to insure that all the materials and information requested are received. If there are any discrepancies, the agency should be contacted immediately, by telephone, to determine if and when the missing materials will arrive.

Each item received, i.e., every element of a product, should be labeled with a product number. This is particularly important with developmental products which are likely to consist of many elements and support documents which do not bear the product's formal title or any form of cross-indexing identification.

In addition, as each product is received, the agency should be notified of its receipt unless the package was sent with a return receipt requested.

A simple card of acknowledgement indicating what materials were received, and when, suffices. Form cards could be prepared in advance, so that only the list of materials and date received need be added.

Because of the size and numbers of products being dealt with, a great deal of storage space will be required. This space should be amply outfitted with cabinets, shelves and tables. This space should also be such that a high degree of security over the materials can be maintained. Not only are the materials themselves expensive, and often attractive, but the ancillary use equipment such as tape recorders, projectors, and the like, are also highly pilferable. The location of each product should be labeled with both the product's title and number to facilitate locating the materials. It is also useful to classify the products by topic area (such that those products to be evaluated by the same panel are stored together) and, within those areas, by agency.

For those products to be returned, it is helpful to save the cartons they arrive in, if the space is available. This greatly facilitates the process of re-packing and shipping the products. If this is done, the packages should be labeled, so that the materials to go in a particular carton can be identified.

DISTRIBUTION OF PRODUCTS FOR EVALUATION

For those products to be mailed out to panelists for evaluation, a system of distributing and monitoring should be established. A return receipt should be routinely requested for all products sent out by the evaluation coordinator to insure that they reach the proper parties. In the case of products being circulated among the various evaluators (rather than each evaluator having his own copy) a follow-up contact should be made at the end of a week (or whatever interval is decided upon) to insure that the products are being forwarded on schedule. In addition, a follow-up should be made on products to be returned to the evaluation coordinator to insure that the evaluators return them.

In the cases of circulating products or products needing to be returned, instructions should be enclosed with the product regarding how the evaluator should dispose of the product when he is finished reviewing it. For products being circulated, a copy of the review schedule and dates should also be included. Finally, address labels for forwarding or returning products should be provided. Postage tallies will need to be maintained in order that evaluators can be reimbursed for their postage fees.

Because of the numbers of materials needing to be sent to the various evaluators, it would be useful to prepare a series of address labels in advance. These can be simple preprinted labels bound into pads with gum backing. They can be used both by the evaluation coordinator and, in the case of circulating products, by the evaluators.

Chapter 9

COORDINATION OF THE EVALUATION EFFORT

Because of the numbers of evaluators, the numbers of products to be evaluated, and the alternative ways in which a given product might be reviewed, careful attention should be given to the procedures for conducting product evaluations. Guidelines for the scheduling and management of evaluations are presented in this section. In addition, specific suggestions regarding each of the three evaluation modes, home/office review, central site review, and field visit review, are included.

SCHEDULING AND MANAGEMENT

The first task in mapping out the evaluation schedule is to determine how each product could best be evaluated. As soon as the initial copies of the products begin to arrive, the evaluation coordinator should review them and assign them to a particular evaluation mode: home/office review, central site review, or field review. The following guidelines should assist him in making these decisions.

1. *Home/Office* review should be utilized if:

   - from seven to ten (depending on the number of evaluators) copies of the product can be economically obtained,
   - one or two copies of the product can be obtained and circulated among the evaluators,
   - the product can be sent through the mail or through some parcel service with relative ease, or
   - the product does not require any elaborate equipment which evaluators are not likely to have access to.

2. *Central site* review should be adopted if:

   - the product is too expensive and awkward to mail,

- a special demonstration of the product will need to be conducted, or
- special equipment will be required to review the product which individual evaluators do not have access to.

3. *Field* review should be utilized if:
   - the product cannot be adequately judged without seeing it in operation, or
   - the product cannot be mailed or shipped but is available for observation at some field site.

Once the preferred evaluation mode for each product has been determined, arrangements should be made for the central site and field observations. For reasons of economy and effectiveness, if possible, the central site reviews should be conducted immediately after the evaluator training sessions, while the evaluators are still together as a group. Depending on the location of the field visits, some or all of these might also be arranged for this time period. In this way the additional costs of reconvening the panel at a later time can be avoided.

In scheduling the products to be reviewed in the home/office mode, those products requiring circulation among evaluators should be considered first in that a larger amount of time will be required for all of the evaluators to receive and review the products. It is suggested that one week be allotted for reviewing a product, and a second week for shipping it to the next evaluator. Allocating one week for reviewing each product allows the evaluators sufficient time to fit the review into their schedules. It is important that a schedule be established and maintained for these products to avoid damaging time delays.

Products for which multiple copies are available should be scheduled in and around those being circulated. However, in the case of multiple copies, all evaluators should review a given product at the same time, so that review

of the results and reconsideration of the evaluations can be completed while the product is relatively fresh in the evaluators' minds. These products might, then, be scheduled for review during the weeks when circulating products are in transit. Again, approximately one week should be alloted for review of each product.

In developing the overall schedule, attention should be given to such factors as holidays, professional society conventions, and so forth, which are likely to affect evaluators' availability. Once the master evaluation schedule has been prepared, it will form the basis for monitoring the progress of the evaluations.

Although it may sound as if an inordinate amount of time is devoted to scheduling of management activities, the significance of the coordinator's contribution in this area cannot be over emphasized. The success or failure of the evaluation effort will be in large measure due to the staff work of the evaluation coordinator in this area.

A second task of the evaluation coordinator is the distribution of evaluation forms and Product Report forms to the panel members. This may be done at the training session, accompanying the schedule mentioned previously, or when the products themselves are distributed. It was found useful during the tryout of the system for the evaluation coordinator to prepare the evaluation forms in advance, filling in the product titles and numbers and evaluator identification numbers. Assigning numbers to use in identifying the evaluators during the evaluation makes it easier to maintain the anonymity of the product evaluations.

In addition to developing a schedule and progress monitoring system, it will be useful for future evaluation efforts for the evaluation coordinator to maintain files on the evaluators. These files might be set up on index cards or perhaps a combination of index cards and support documents. Whatever the form, the files should include basic information about the evaluator, such as his name, identification number, address (both residence and business), tele-

phone number(s), biographical references, and fields of specialization. To this information should be added records on his function as an evaluator, such as who nominated him, what products he reviewed, what his expenses were (e.g., travel, honoraria, and per diem) and comments by the evaluation coordinator on his general performance as an evaluator, his points of view, his availability to serve in the future, the nature of his contributions, and so forth.

A second file, a suspense file, should be established for storing completed evaluation forms as they are received. Forms should be kept by the evaluation coordinator in case any of the product developers file exceptions reports or request backup evaluations, or in case any of the evaluators file minority reports. However, the forms should not be held longer than six months after completion of the evaluation, in order that the file may be purged prior to the implementation of the system in the following year. In this way, the accumulation of confidential data will be precluded.

The remainder of this section will present specific suggestions for the conduct of the three evaluation modes. Because of the different conditions and demands of the three modes on the evaluation coordinator, generalizations across the three modes regarding his responsibilities cannot be made.

HOME/OFFICE REVIEWS

It is likely that the majority of products will be reviewed in evaluators' homes or offices. The evaluation coordinator is responsible for insuring that the system functions smoothly. Thus, the coordinator's task will be greater in this mode where he has nine individuals to keep track of rather than one group of people.

Before the first products for review are sent out the evaluators should be briefed on what will be expected of them. This may be done either at the conclusion of the training session, if the training immediately precedes the home/office review, or through a mailed package of information, followed up by a

conference call. The evaluators should be given a copy of the review schedule indicating what products they will be receiving, when they should plan to review each, what should be done with each when they have completed their review, and what equipment, if any, will be required for the review. In addition, they should be given address labels for those products requiring either forwarding or return. If the evaluation forms are distributed at this time, self-addressed, stamped, return envelopes should be enclosed for each form.

Once the evaluators have begun reviewing the materials, periodic telephone contact should be maintained to monitor their progress. This is particularly important in the case of products being circulated among evaluators to insure that a product does not get hung up on one evaluator's desk, throwing the schedule off for the other evaluators.

If an evaluator requests additional information about a particular product, the evaluation coordinator should prepare a standard reply to the request and send it to all the evaluators. It is particularly important when the evaluators are not together in a group that all evaluators receive the same information.

FIELD REVIEWS

In those instances in which the evaluators travel to field sites, either individually or as a group, the evaluation coordinator will be responsible for arranging the visits. As soon as the location and tentative dates have been identified, he should contact the responsible staff member at the site and confirm a date and time when the observation can occur. At this time, he should apprise the staff member of the purpose and objectives of the visit. He should emphasize that evaluators be given an objective view of the product and be allowed to observe and examine all relevant elements of the product.

The coordinator should also indicate to the local staff what is <u>not</u> wanted. Site staff may be tempted to talk about the "potential of the product" instead of "what it is"; about how "well" it operates rather than "how" it operates, etc.

This, of course, should be tactfully avoided. The tryout of the evaluation system included one presentation by the developing agency. By carefully explaining what kinds of information the developer should cover and what kinds to avoid, a reasonably direct and informative presentation resulted.

Shortly before the field visit is to occur, the evaluation forms and any support materials should be sent to the evaluators. (In the case of a central field visit, forms and materials can be distributed when the evaluators convene.) Although the evaluators will have previously received a schedule indicating the time and place of the visit, they should be reminded of the arrangements at this time.

If the product can be seen at many field sites, the coordinator may wish to make arrangements for viewing the product at the sites most convenient for individual evaluators. The evaluation coordinator will still, however, be responsible for briefing the local site staffs on the purpose of the observations.

Whenever possible, both group and individual field visits should be supervised by the evaluation coordinator or one of his staff. This is particularly important with group visits in which the evaluators will be tempted to discuss the product they are observing. In order to preserve the independence of evaluators' initial ratings, it is necessary to avoid such discussions.

CENTRAL SITE REVIEWS

Several of the suggestions regarding the field review mode of evaluation will also be relevant here. Of most importance is the presence of the evaluation coordinator or one of his staff to make sure that unwanted discussions do not occur.

The evaluators should review a product and complete the evaluation form before moving on to the next product.

Occasions in which discussion is likely to arise, such as during lunch or breaks, should be scheduled to occur after evaluators have completed their review and rating of one product, and before starting the next.

Similarly, if a special demonstration of a product is to be held, as in the case of the field visits, the demonstrator should be cautioned to provide only an objective description of what the product is and how it works.

If several products are to be reviewed in the central review mode, a member of the evaluation coordinator's staff should review the products prior to convening the panel in order to determine the approximate amounts of time which will be required for the evaluators to examine the materials and make their decisions.

During the system tryout several of the evaluators felt that they were not allowed sufficient time to review a product; thus it is probably better to err in over-estimating the amount of time required to review products.

If there are numerous materials associated with a particular product, an element rotation schedule should be devised, so that some evaluators needn't wait until the other has completely finished examining the product.

Separate rooms should be made available for evaluator use, both to provide an environment conducive to materials review and to minimize the possibility of inter-evaluator discussion of the materials.

If special equipment is to be used for a demonstration or review of a product, the evaluation coordinator should obtain and check the equipment prior to the time it will be needed. In the tryout of the system, it was necessary to rent a broadcast quality videotape recorder to play video tapes. Although the machine received was the model requested, and it was supplied by a highly reputable television company, it required some adjustments by a technician in order to obtain clear reception.

In order not to lose time waiting for the evaluators to convene, they should arrive on the evening prior to the first day of the meeting. In this way they will all be present and can begin their tasks promptly in the morning.

Depending on the duration of the group meeting, it may be desirable to include some social activity in the agenda. For example, if the session is scheduled to last two days, a no-host cocktail hour and dinner might be planned for the evening of the first day. This provides an excellent opportunity for the evaluators to discuss products outside the evaluation context after having been forbidden to discuss them during the day. It also gives the coordinator the opportunity to evaluate the performance of the panel members and to ask questions regarding the evaluation procedure which may result in the eventual improvement of the procedure.

PROCESSING THE RESULTS OF THE EVALUATION

The third main responsibility of the evaluation coordinator is to process the results obtained. This involves circulating the initial product ratings, analyzing the final ratings, and preparing the evaluation panel reports. Each of these is discussed in the following sections.

Distribution of Initial Ratings. When all the evaluators' ratings for a specific product have been received, the evaluation coordinator should circulate the results among the panel members, asking them to reconsider their ratings in light of the other evaluators' judgments and to modify them if they see fit.

Xeroxing the individual rating forms, minus any evaluator identification is the most efficient method of distributing the results. The evaluators' comments, as well as their ratings, can thus be considered.

During the system tryout, ratings were recorded on summary sheets along with abstracted comments. While the mechanics of this type of distribution were simpler, being based on nine one-page sheets rather than nine eight-page forms, this approach was felt to be less useful in that it was impossible to fully convey all the flavor of all of the evaluator comments.

The evaluation coordinator should reproduce and distribute the ratings for a product as soon as possible after they are received. In this way the product should be relatively fresh in the evaluators' minds when they reconsider their ratings. In the case of products which were circulated among evaluators, such that many weeks may have lapsed between their initial examination and the receipt of the initial ratings, the evaluation coordinator should suggest that the evaluators review the Product Report Form to refresh their memories about the product. The evaluator's original rating form should also be returned along with the copies of the other rating forms, in case he wishes to modify any of his earlier judgments. A stamped, self-addressed envelope should also be enclosed for returning the evaluator's original form.

Negotiation of Final Ratings. When the revised ratings are received, they should be transcribed to a Rating Summary Sheet, along with the more critical comments. An example of such a form is provided in Appendix D. In those instances where there is a discrepancy of more than one point for more than one evaluator, the evaluators should be asked to discuss, jointly, the arguments underlying their respective decisions.

If the evaluators are together, in the case of a central site review or a group field visit, then the discussion can be conducted at that time. In the case of mailed products, where results are sent in, the discussion can be conducted via a telephone conference call. An average 20 minute, 10 station conference call will cost approximately $60.

If the products to be discussed have been circulated among the panel members, such that many weeks may have passed since the first reviewers examined the product, the evaluator should alert the evaluators that such a discussion will take place in approximately a week to allow them to refresh their memories regarding the product.

The evaluation coordinator may find it helpful to contact each of the evaluators by letter prior to the conference call to confirm the date and time of the call, the product or products to be discussed, the ground rules for the discussion, and to advise the panel of the range of ratings on the criteria in question.

The evaluation coordinator should participate in the conference call in order to guide the discussion. He should identify the criteria in question, review the distribution of ratings on these criteria, and query the evaluators regarding their reasons for particular ratings. By focusing the discussion on the issues of concern, he will avoid wasting time.

If evaluators wish to modify their judgments at this stage, they may still do so. For those criteria for which the variance is not resolved, however, the evaluation coordinator should note the reasons for the variance so that they can be indicated in the discussion of results in the evaluation panel report.

Processing the Data. When the final ratings have been compiled, the mean ratings on each criterion for each product should be calculated and recorded. Once the mean ratings have been determined, the evaluation profiles should be plotted on an Evaluation Summary Sheet. In processing the data obtained from the tryout of the system, it was found that graphic profiles provided the most meaningful display of the evaluation data.

Several different types of data displays can be prepared, as exhibited in Appendix C. The Evaluation Summary Sheet depicts the profile of a specific product in relation to the profiles of the other products of the same type (knowledge or developmental) with which it was reviewed. The Scatter Plot simply shows the variance of mean ratings across all knowledge or developmental products on each of the criteria. These profiles, in conjunction with written comments, form the base of information on which the evaluation panel reports are prepared.

In preparing these profiles, it was found useful to indicate the "average" range, defined as approximately the middle third of the scale. By using this band (which covers approximately $\pm$ .6 SD) and the corresponding above-and-below average bands ($> \pm 1$ SD), it is easy to identify which products tend to receive average, above-average, or below-average ratings on the various criteria. The portrayal of these ranges, however, is intended only as an heuristic for interpreting the data. For this reason, band widths of .4 points, rather than single lines, have been used to delineate the three ranges in order to emphasize the arbitrariness of conclusions regarding "borderline" ratings.

The evaluation coordinator must be continually sensitive to the fact that it is not his function to make public distribution of the results of particular product evaluations. Thus, the coordinator should be careful when providing evaluator feedback to agencies to mask the identities of all products except those they personally developed. In order to provide a meaningful framework for the interpretation of evaluation results, however, the distribution of ratings of all similar products is necessary. The Evaluation Summary Sheets serve this function without compromise of the identity of others' products. In this way the anonymity of the results is preserved, but a frame of reference for interpreting the results is provided.

Similarly, in preparing the Rating Summary Sheets the evaluation coordinator should take care that the individual evaluators are not identified. Assigning an identification number or code to each evaluator, as mentioned earlier, will obviate this difficulty.

REPORTING RESULTS

Upon completion of the evaluation effort, the evaluation coordinator should prepare a report on the activities and findings of each evaluation panel, plus an overall summary evaluation report. These reports are not intended for general distribution but, rather, for use by USOE or NIE program planners.

Each of the reports should follow approximately the same format. Basic information on the panel activities should be provided, including the dates and settings in which the evaluations were conducted, the products evaluated, and a brief statement of the background of each of the evaluators.

In addition, any special conditions prevailing during the evaluaton which may have implication for interpreting the results should be documented. This would include reasons why a given product could not be evaluated as intended or why deviations from the recommended evaluation mode occurred.

Finally, the results of the evaluations should be discussed. The Multiple Profiles Sheets for the products reviewed and the individual product Evaluation

Summary and Rating Summary Sheets should be included to provide the basic information on the results obtained. Discussion should highlight the findings, indicating for each product any areas where ratings tended to fall in the above- or below-average ranges, or any trends occurring across products. For example, in the system tryout it was found that ratings on content clarity and accuracy tended to be generally high across all the developmental products; this pattern was pointed out in the discussion of results. It is also important that any qualification of the results be specified. Samples of various data summarization sheets resulting from the pilot test, with all product identities removed, are presented in Appendix D.

The Summary Evaluation Report, as its title suggests, summarizes evaluation results across all the panels. Basic information covered should include the number of products reviewed and the topic areas dealt with, the composition of the evaluation panels, and, in general, the settings in which the evaluations occurred. The various evaluation panel reports serve as back-up material for this summary report.

PART III

PILOT TEST RESULTS
AND
RECOMMENDATIONS FOR FUTURE IMPLEMENTATION

Chapter 10

PILOT TEST RESULTS

Two separate evaluation efforts were carried out in the pilot test. These efforts were conducted solely for the information they would afford toward the improvement of the evaluation system. For this reason, the evaluation panel meetings were operated so as to maximize opportunities for obtaining useful feedback. In so doing, some compromise of evaluator independence was, of course, necessary. Thus, the evaluation effort discussed herein should be viewed primarily as a simulation of the recommended process.

The primary difference from the recommended evaluation model lies in the fact that all product evaluations were conducted at a central site, i.e., the evaluation coordinator's office. Normally most products, an estimated average of approximately 80%, would be distributed to evaluators for review in their homes and/or offices. In the interest of holding discussions about the strengths and weaknesses of the system, however, as well as maintaining a close check on its operation, the evaluation was conducted in a central conference mode. Two products were evaluated under simulated mail conditions, though, i.e., under conditions where evaluators reviewed products in the leisure of their own homes. Further, even though all evaluators were physically present at AIR, an attempt was made to maintain the independence of evaluators' judgments by assigning each evaluator to a private office where he reviewed and evaluated products and by prohibiting the mutual discussion of products prior to, or during, their evaluation.

SPECIAL FACTORS

As many exceptional cases as possible were incorporated in the tryout. The purpose was to test the system's limits, to test its applicability under stress. One product required a special field visit to a neighboring city to see the product in operation in a neutral setting (in a setting where the product developer was not present, but the product was in use). In several other instances, special audio-visual equipment was necessary; and in another

instance where a field visit to view the product was not feasible, the product developer accompanied the product to the conference site and gave the evaluators a brief verbal orientation to the product and its complexity. Upon completion of his presentation, he left the area so that there would be no further influence on evaluator deliberations.

Several other special factors were also introduced into these sessions. Some evaluators were local while others traveled great distances; also, some evaluators represented users whereas others represented researchers, product developers, and evaluators. In one instance the same product was evaluated by two different panels by virtue of the fact that the product was extremely complex and, as a result, was jointly classified under two different headings in the product taxonomy. Finally, one product was evaluated against criteria that seemed somewhat less than appropriate in that the product developer reported the product as a developmental product, and persisted in doing so in a follow-up check, even though it seemed to the panelists more appropriate to consider it a knowledge product.

Another major area of concern had to do with individual differences in the reading speeds of various panel members. Under the tightly controlled time constraint of the conference mode of operation, it was necessary to assign fixed periods of time for the review of each product. For some evaluators, the allocated time was more than ample; for others, the time was too short.

Finally, as a concession to the subsequent critique of the system by the panel, the primary purpose of the tryout, panel membership was held to only five panelists, as contrasted to the six or eight which would normally constitute a full evaluation complement. Inasmuch as three project staff members were integral to the panel, and, in one case, there were OE visitors as well, it was necessary to keep the total number of the aggregate group on the order of eight to ten so that candid interaction of the group could be facilitated in the critique of the evaluation system.

PRODUCT SELECTION AND EVALUATION

The products reviewed by the evaluation panels were selected from those products reported as completed by the R&D Centers and Regional Educational Laboratories.

The selection of products was based on their taxonomic classification. As there were insufficient numbers of products in single classification cells to warrant convening disparate evaluation panels, related categories were clustered together so that the resulting group represented 10-12 products. Two of these clusters, containing a total of 22 products, were then selected for review. The clusters were: The Learner and the Learning Process, and the Design and Development of Educational Systems.

It should be apparent that the sample of products reviewed by the two panels of evaluators is not, nor was it intended to be, representative of the entire domain of educational products, or even of all products produced by Regional Laboratories or R&D Centers.

When the 22 products were requested from their respective developers, the evaluation coordinator was advised that two of the products were no longer available for evaluation. In one instance, the agency declined to provide the product, asserting that it was of only minor importance and not developed as part of a formal agency program. In the second instance, an item reported as a product of a laboratory turned out to be conceived, funded, and developed by an independent concern and was thus solely proprietary to that concern.

Further, of the 20 products remaining for evaluation, it was found that two appeared to have been completed prior to the establishment of the reporting agency but were reported as accomplishments by virtue of the fact that the author had subsequently become a staff member by the time the products were published.

Thus, it would appear that approximately 18% of the products reported as having been completed by laboratories and centers have some question attached to them.

These 20 products were evaluated by two product evaluation panels during the early weeks of May, 1972. One panel evaluated products dealing with characteristics of the Learner and the Learning Process; the other panel evaluated products concerned with new Instructional Systems. Each evaluation session lasted two and one-half days. During that time each panelist reviewed approximately ten products.

DEVELOPMENTAL PRODUCTS RESULTS

Eleven of the 20 products evaluated were developmental products. All told, members of the evaluation panels made over 600 separate, individual judgments. Individual ratings of specific products on a given criterion were then averaged across evaluators to yield a "panel judgment" on that criterion. This resulted in a total of 115 separate panel judgments. Thirty-eight percent of the panel judgments were in the "above average" category. Only 11% were judged "below average." Thus, some 89% of panel judgments regarding developmental products were average or above average.

Of the 11 developmental products evaluated, five received consistently high ratings, that is, five accounted for the bulk of all above average ratings. Three products accounted for all "below average" ratings. One of the three, however, received below average ratings only in regard to its amenability to marketing and potential impact. Otherwise, it was judged to be in the average range for products of its type.

One product was evaluated by both panels. It is interesting to note that the evaluation profiles produced by the two independent panels are highly similar. See Figure 15. This suggests a fairly stable and reasonably valid evaluation even though the two panels were quite different in composition, and the form of product description to the panels differed considerably. In the first panel, the product developer made a brief presentation with videotape demonstrations. In the other instance, no videotape playback was used and no special presentation was made other than a brief factual description by the evaluation coordinator.

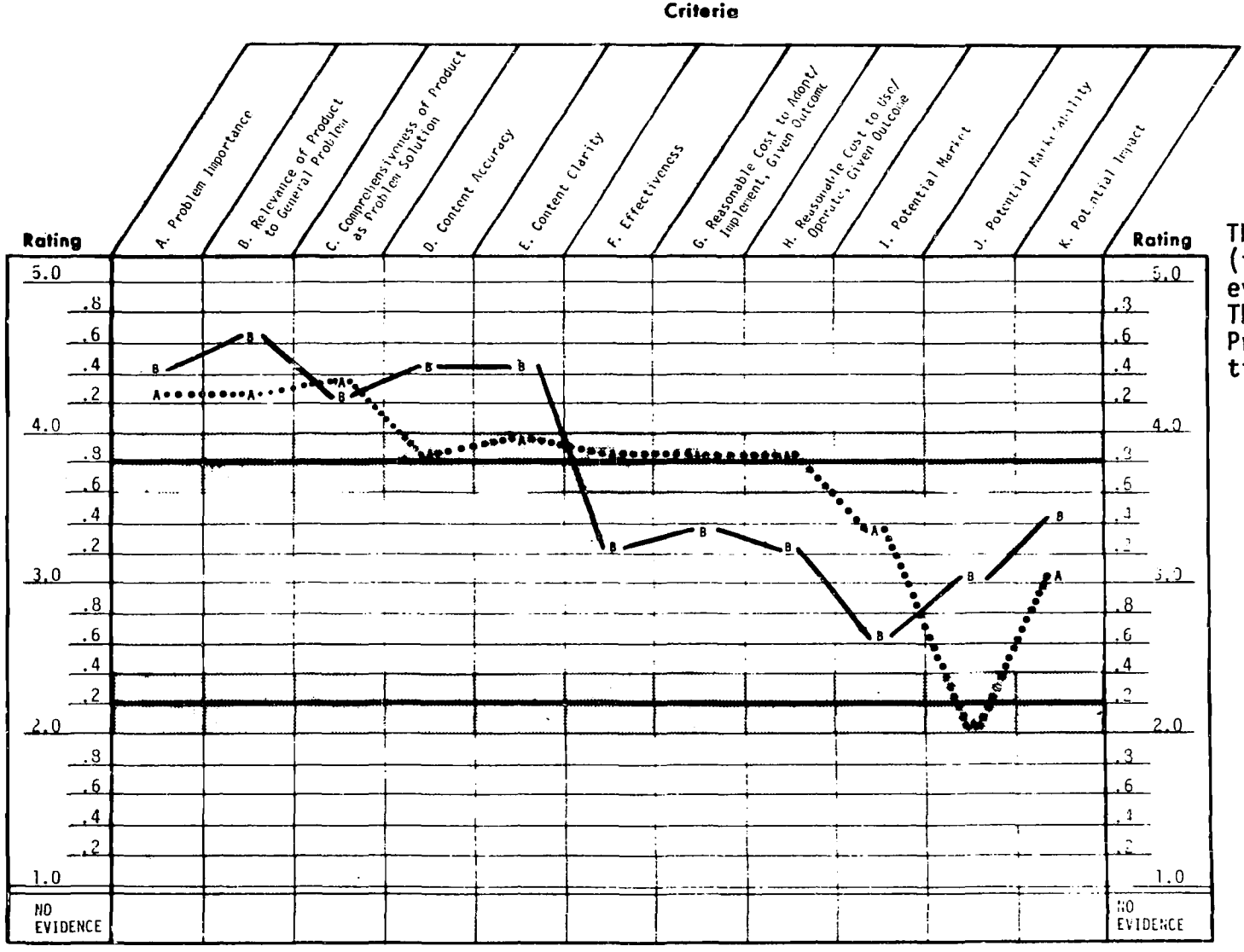
DEVELOPMENTAL PRODUCTS MULTIPLE PROFILES SHEET

DATE 3 May & 10 May 1972

PRODUCT AREA The Learner and the Learning

NUMBER OF EVALUATORS 5 per Area

Process: Educational Systems



The ratings of your product (identified above) by two evaluation panels: Panel 1, The Learner and the Learning Process; and Panel 2, Educational Systems

PRODUCT CODE

A Panel 2

B Panel 1

C

D

E

F

G

H

I

J

Figure 15

EVALUATION OF THE SAME PRODUCT BY TWO DIFFERENT PANELS

It is especially important to note that support documents were submitted for only three of the products. There was an almost total absence of effectiveness data submitted by developers for the panels to consider in judging the effectiveness of their products. It is not known whether this is because no evaluation data had been collected; or whether they had been collected but were not yet analyzed or written-up sufficiently to warrant submission with the product; or whether such data had been collected and the evidence was non-supportive.

This lack of empirical evidence of the effectiveness of the completed products is quite typical, however, of most products on the market and this may be the reason why panelist's judgments of the effectiveness of the products tended to cluster very closely around the center of the rating scale, i.e., 3.0. Inasmuch as no evidence was submitted in support of the products, the evaluators had only developers' assertions of their product's effectiveness; and this was quite typical of products in general. As a result, just as there was no evidence for rating the product above average, there was similarly no contra-indicative evidence which would result in rating the product below average on the effectiveness criterion.

KNOWLEDGE PRODUCTS RESULTS

With regard to knowledge products, a total of 97 separate panel judgments[1] were made on 10 products. Thirteen percent of the evaluator judgments were "above average," 66% were "average," and 21% were "below average." The bulk of the above average ratings were contributed by one set of reports. The bulk of the below average ratings were contributed by three single-study products. It is interesting to note that two of these latter three knowledge products were published only as in-house reports and filed with ERIC. They were not published in refereed journals or by commercial publishers.

---

[1] Based on over 450 panelist ratings.

COMPARISONS OF RATINGS ACROSS PANELS

It is also interesting to note that for developmental products, the range of judged potential impact of the educational systems products and the learner products is essentially the same, but the judgment of problem importance is considerably lower for educational systems products than for the learner products.

The range of judgments on content accuracy and content clarity is also essentially the same for both product groups, as is the range of reasonableness of cost, thus suggesting comparable levels of craftsmanship.

Judgments of the scope of the possible market for learner oriented products are considerably higher than for educational systems products. The higher potential market and the higher judgment of problem importance for learner products, as compared to educational systems products, may suggest differences in the basic missions of the two groups.

Regarding knowledge products compared across learning and educational systems, the learner products were judged much higher in importance than the educational systems products. There was much greater relevance of knowledge products to the problem area for educational systems than for the learning area. This may be a function of inflated rhetoric in the problem statements of the learning group.

The comprehensiveness of knowledge products as a problem solution seems to be somewhat greater for the educational systems group. The range of originality of knowledge products is about the same for both groups. The adequacy of research design tended to be considerably higher for the learning group than for the educational systems group. There appeared to be no real differences, though, in the reasonableness of conclusions, the clarity of presentation, or the judged potential impact of the two groups.[1]

[1] It should be remembered that these statements are based on interpretation of the data from only two small sets of products. Data from a considerably larger number of products would be necessary before such generalizations can be taken for anything other than their heuristic value. They do, however, suggest directions that may be pursued when a sufficient number of products has been evaluated. Additional types of questions that may be asked of the product information/evaluation data base are discussed in Chapter 4.

RATING SCALE CHARACTERISTICS

Figures 16 and 17 show scatter plots of the evaluation judgments for both panels combined. (All ratings made using the two stage, seven-point, scale have been converted to five point equivalencies.) It would appear from inspecting these figures that the procedures used did effectively differentiate products on the various criteria.

If all 212 panel judgments in the pilot test are pooled and analyzed statistically, a mean rating of 3.05 and a standard deviation of 1.01 is obtained. Thus, the evaluation procedures in general result in a distribution of scores centered on the mid-point of the rating scale with a standard deviation of approximately 1 rating scale point. There is, of course, variation in these values depending on the criterion and the type of product being considered.

Given developmental products the mean ratings on the 11 separate criteria range from 2.5 to 3.9, with a mean of 3.29. The standard deviations of the ratings on the 11 criteria range from .72 to 1.15 with a mean standard deviation of 1.02.

For knowledge products mean ratings on the 10 criteria range from 1.9 to 3.5 with a mean of 2.83. The standard deviations of ratings on the 10 criteria range from .82 to 1.12, with a mean standard deviation of 1.01.

The evaluation procedures proposed in this study, then, appear to result in quantitative judgments of products which afford considerable convenience in statistical interpretation.

The scales also manifest a reasonable degree of construct validity. Ratings on the various criteria were intercorrelated and then subjected to two forms of "cluster" analysis: elementary linkage analysis (McQuitty, 1957), and principle components normalized verimax factor analysis. The results from both types of analyses were essentially identical.

SCATTER PLOT -- ALL DEVELOPMENTAL PRODUCT EVALUATIONS COMBINED

DATE 3 and 10 May 1972

PRODUCT AREA The Learner and the Learning

NUMBER OF EVALUATORS 5 per Area

Process: Educational Systems

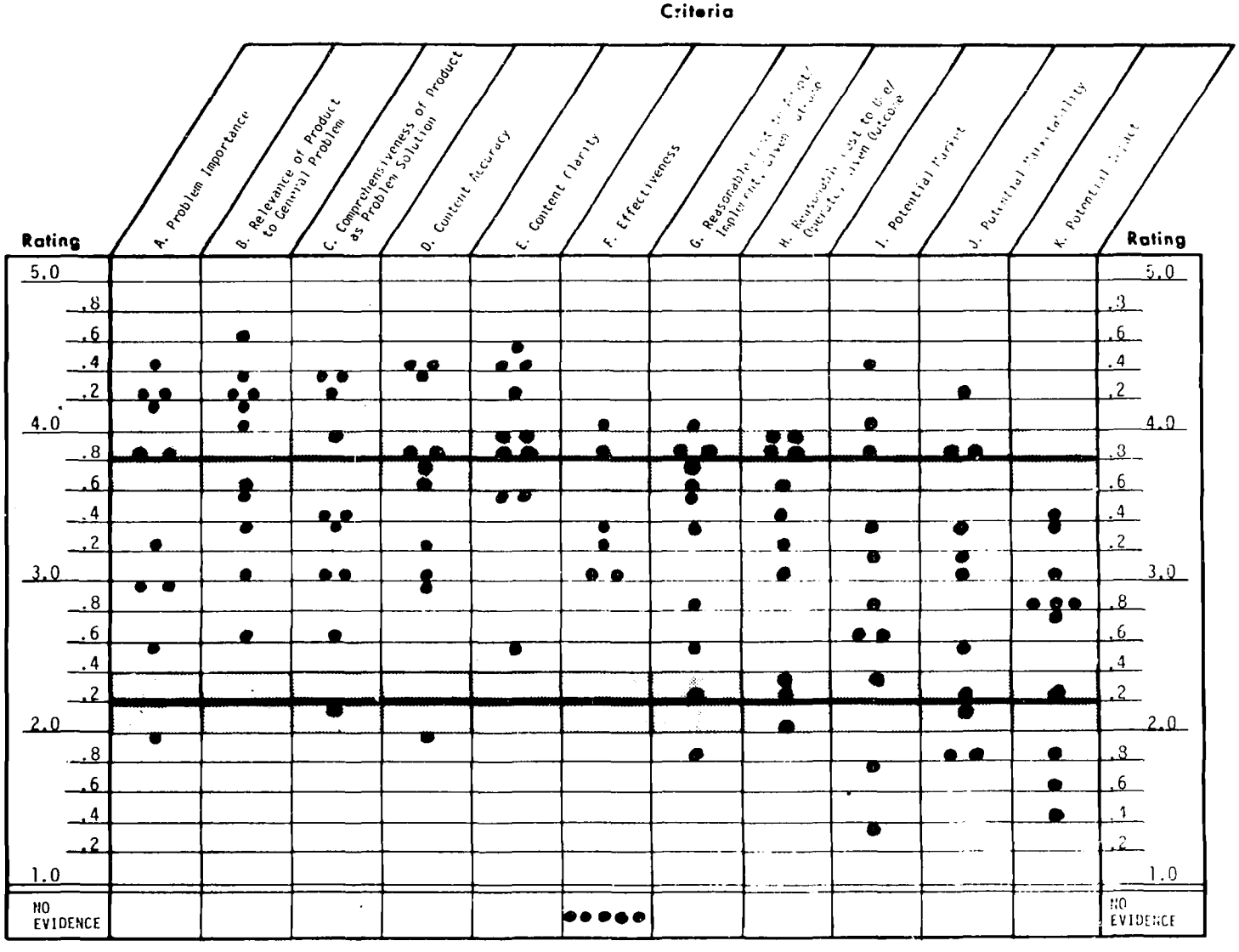


Figure 16

SCATTER PLOT -- ALL KNOWLEDGE PRODUCT EVALUATIONS COMBINED

DATE 3 and 10 May 1972

PRODUCT AREA The Learner and the Learning

NUMBER OF EVALUATORS 5 per Area

Process: Educational Systems

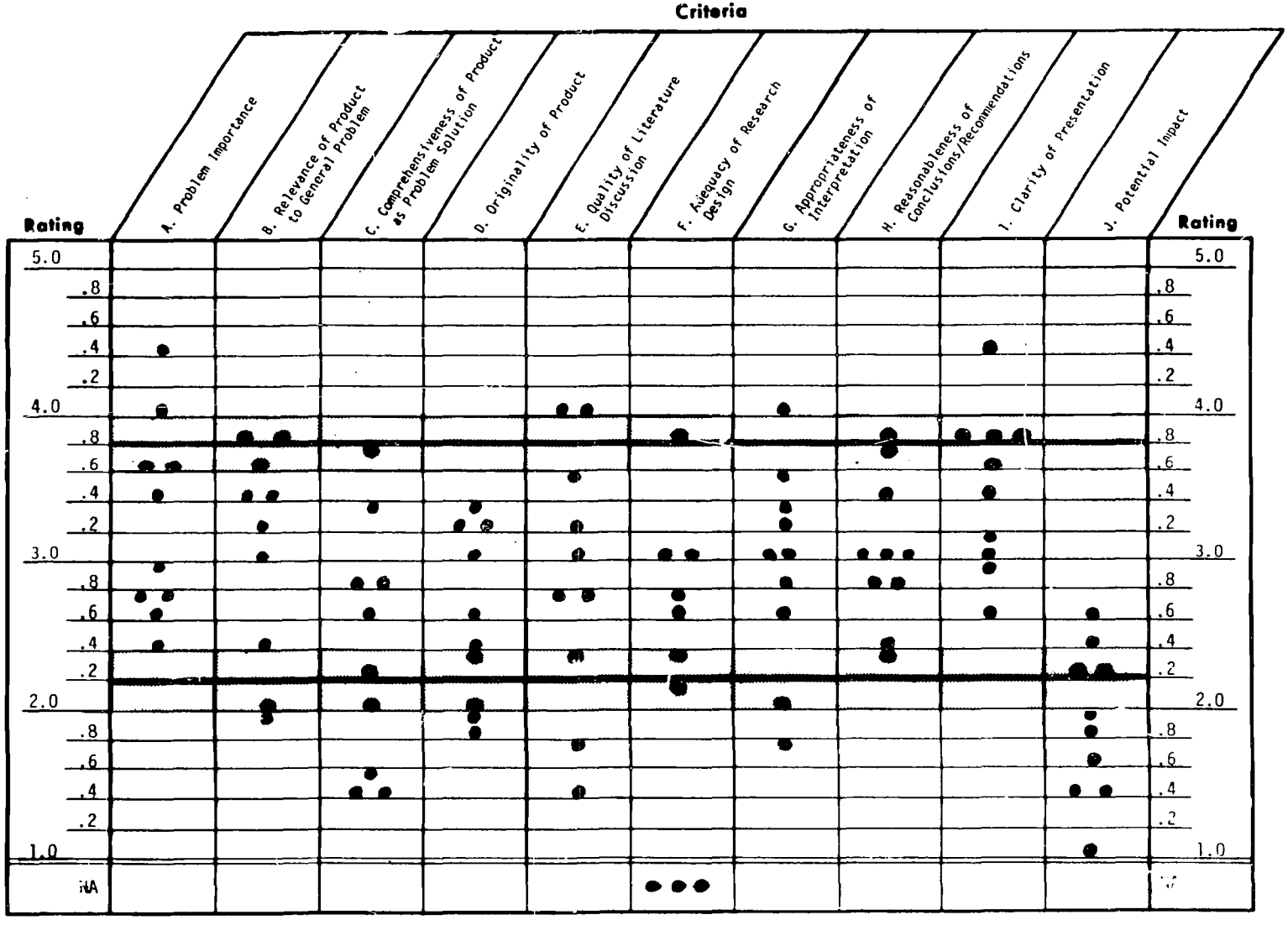


Figure 17

The results of the McQuitty's elementary linkage analysis are summarized in Figures 18 and 19.

In the factor analysis of ratings on the 11 criteria for developmental products, 4 factors accounted for 87% of the common variance. Factor A was labeled Product Significance. This factor included ratings on problem importance, potential impact, and scope of possible market. Factor B was labeled Quality and was defined by the criteria of content accuracy and content clarity. Factor C was defined by the criteria of effectiveness, comprehensiveness of the product as a problem solution, and relevance of the product to the general problem. Factor D was defined as Practicality. Products high on this last factor were judged to be attractive, easy to use, and of reasonable economic cost to adopt and use, given anticipated outcomes.

Four factors accounted for 82% of the common variance in the product evaluation judgments on the 10 criteria for knowledge products. Factor A was labeled Significance. Products high on this factor would be judged to be important, and to be carried out in a highly competent manner. They would manifest good research design, embody a good literature discussion, appropriate interpretations of the data, and reasonable conclusions and recommendations based on those data. Factor B was labeled Quality. Products high on this factor would be judged original, comprehensive, and of high potential impact. Factor C was a stylistic factor which was defined by the single criterion, Clarity. Factor D was also a single item factor defined by relevance of the product to the general problem.

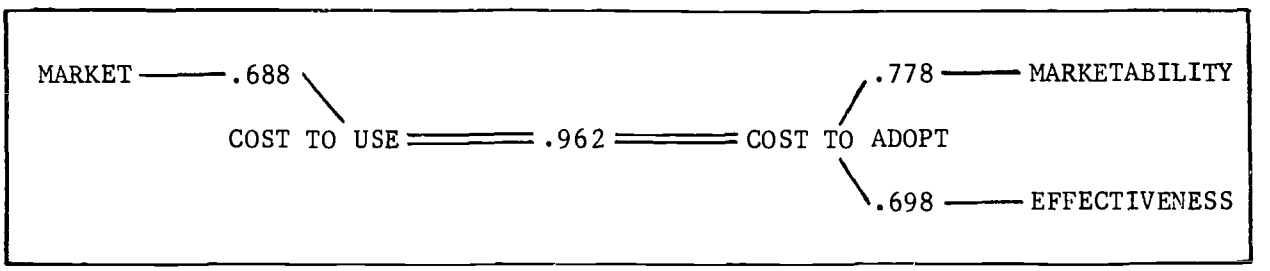
Item communalities for the developmental products ranged from .81 to .96 with a mean of .87. Item communalities for the knowledge products ranged from .68 to .93 with a mean of .82. Since item communalities represent only common factor variance, and since the true score variance of an item is composed of the sum of common factor variance and specific factor variance, item communalities constitute a lower bound, i.e., maximally conservative, estimate of item reliability. Thus it would seem that the procedures developed for this evaluation system result in panel judgments of considerable reliability.

Figure 18

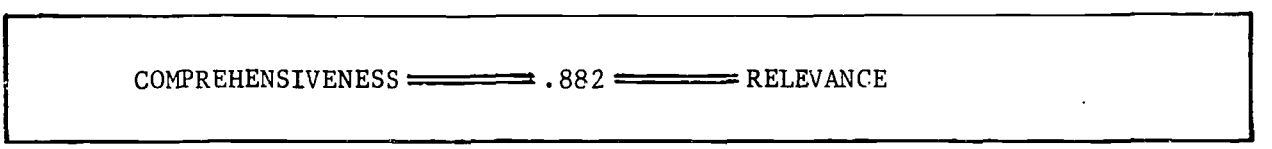
ELEMENTARY LINKAGE ANALYSIS

Developmental Products

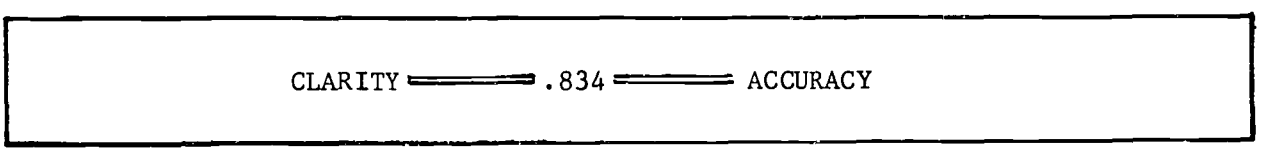
Cluster 1



Cluster 2



Cluster 3



Cluster 4

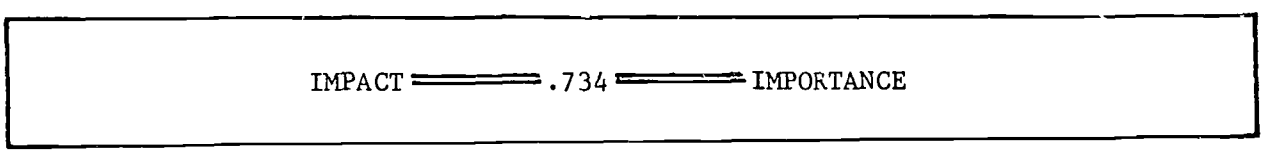
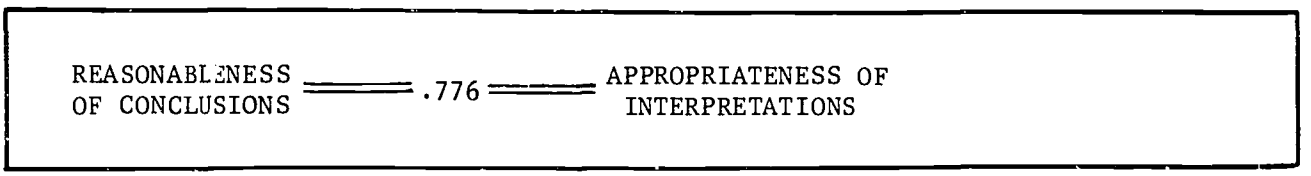


Figure 19

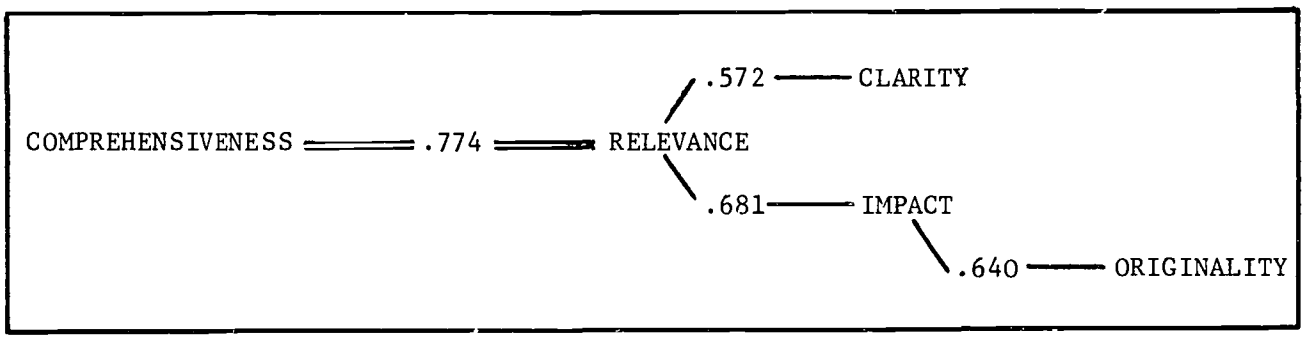
ELEMENTARY LINKAGE ANALYSIS

Knowledge Products

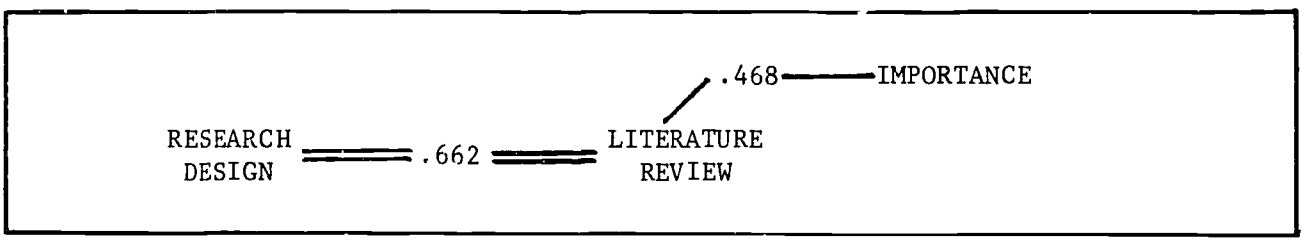
Cluster 1



Cluster 2



Cluster 3



Finally, the data suggest that the two stage, successive judgments method might be a more effective method of judgment than the single stage method. The standard deviation of judgments produced using the two stage model exceeded the standard deviation of judgments using the one stage model in 86% of the cases. Too much credence should not be given this finding at this stage, however, inasmuch as it is impossible to determine whether this effect was due to the rating methodology itself, or to differences in the individuals using the various instruments.

Chapter 11

RECOMMENDATIONS FOR FUTURE IMPLEMENTATION

Based upon experience gained from the pilot test of the evaluation system, a number of recommendations for system revision and future implementation can be made. Periodic modification was, of course, incorporated into the system during tryout. Some suggestions for revision, however, are the result of the final stages of pilot testing and must, of necessity, await future incorporation should a decision be made to implement product evaluation.

Final recommendations for the system fall into two categories: suggested revisions in product reporting procedures, criteria, and instrumentation; and cost projections for operation of the system in alternative configurations.

SUGGESTED REVISIONS

Product reporting. Without doubt the most difficult problem encountered in this effort was identification of the product outputs of laboratories and centers. It would seem that this should be a relatively straightforward task. In point of fact, it was not as simple as it might seem. To the extent that the quality of one's output in the past can be construed as an index to the likely quality of one's output in the future, it is understandable that some developers might be quite hesitant to have their products rank ordered for inspection.

Given the anticipated funding policy of NIE, however, comprehensive reporting of all laboratory and center products in the future may be only an academic question. Nevertheless, many of the following recommendations would still be valid regardless of the scope of product reporting involved. The following are the revisions recommended for product reporting.

Product reporting should be made an explicit requirement of laboratories and centers. As such, specific tasks should be written into annual scopes of work. (In view of the length and detail typically reported in the pilot study, if an agency has kept adequate records on its product development, it should require no more than two to three man hours, plus perhaps an additional man-hour for typing, proofreading and clerical review, per product.)

In order to minimize error on the part of the recipient who monitors the influx of reports, agencies should aggregate their product reports and submit all reports from their agency at a single point in time. That is, reports should not be submitted piecemeal.

Product updating should be on an exceptions basis. That is, when new product reporting is carried out, reference should be made to the earlier report on the product (e.g., the "in-process" report); and only relevant section entries should be updated. In this connection the product reporting form should be revised to make provision for the agency director to reference an earlier report on the same project.

For example, upon reporting product X as having been completed, the form should make provision for calling attention to the fact that product X was reported previously on such-and-such a date; and, if the title of the product has changed, indicate the title of the product as it was previously reported.

As a procedure to urge product developers to specify support documents for evaluation consideration, an area should be included on the form where the product developer is asked to specifically cite all support documentation he would like considered in the evaluation of his product. Developers should be informed that lack of citation of field test data will be interpreted as zero field test data.

A reinforced emphasis should be made on the definition of a knowledge product as a contribution of new knowledge made available to the professions through regular publication.

Completed knowledge products should be divided into two groups, those products that are typically available through standard library services, such as books and journal articles, and those that are available through nonrefereed, indexed, "fugitive document" retrieval channels such as ERIC.

Published books will, on occasion, present a problem. On the one hand they are "commercial products" in the sense that they are revenue producing. Most books, however, would not qualify as a contribution of new knowledge to the profession so there should be no problem. The majority of books fall in the categories of instructions to practitioners, guides on how to employ new techniques already developed, or overviews of an area already mastered by most experts in that area. Textbooks, for example, or books on computer programming, basic psychology, teacher training, and the like, would be classified as developmental products.

On the other hand, some books, which are also revenue producing, report major new breakthroughs in science and technology, and, thus, would qualify as knowledge products. These are usually reports of major research programs, however, and will be relatively infrequent.

Evaluation and feasibility studies, while technically knowledge generating, are of extremely limited use and, in most cases, would be submitted as support documents for developmental products. In some cases, however, evaluation studies are of major public and professional interest, such as evaluations of the national Head Start program, or the Follow Through program, or some other major educational endeavor. In such cases they would constitute a source of significant new knowledge regarding a problem of major interest to education.

Section 5 of the knowledge product reporting form should be revised to allow the author to report not only the number of associated publications

that should be aggregated into the composite knowledge product, but also to indicate the serial position of the publication in hand. For example, instead of simply indicating there are five other products dealing with the same general problem area, the author should indicate that, for instance, the publication he is reporting is the third in a series of five publication studies dealing with the same general problem area.

Provision should also be made for him to indicate the level of development of each study in the series.

Finally, the knowledge product form should be revised to allow the author to report all of the other ancillary publications that resulted from the study. It is necessary to clearly indicate to the respondent that he should report the single most comprehensive treatment of the general problem area that he has and that is what constitutes the "product." The variety of miscellaneous publications that may derive from a single research program may be quite large. A half-dozen separate publications may be generated by a single study. It is not desirable for a different report to be filed on each and every ancillary publication deriving from each research study.

Criteria. The intercorrelation of evaluations across criteria, coupled with a survey of the questions evaluators asked during the field test, suggests that the evaluation procedure can be tightened up somewhat. Several criteria can be combined, or eliminated, without apparent loss. By so doing, the work of evaluators can be reduced and the task of data interpretation can be made easier. For example, there is a high correlation (.96) between developmental products criteria "reasonableness of cost to adopt, given outcome," and "reasonableness of cost to use/operate, given outcome." These criteria should be rewritten as a single cost criterion.

Similarly there is a high relationship (.88 for developmental products and .77 for knowledge products) between the criteria "relevance of the product to the problem" and "comprehensiveness of the product as a solution to the problem." While it is theoretically possible to have a product that is relevant but not comprehensive, it is not possible for a product to be a comprehensive solution to a problem yet at the same time be irrelevant to that problem. In view of the empirical evidence, however, it does not seem reasonable to continue to carry the relevancy criterion for the few times it may be appropriately applied. Therefore it is suggested the relevancy criterion be dropped.

These two changes will result in 9 criteria for developmental products and 8 for knowledge products.

Instruments. A number of minor modifications of the instruments and the instruction manuals to accompany those instruments should be made.

If a decision is made to reduce the number of criteria and to redefine others, as in the case of the redefinition of the "cost to adopt and use" criteria, instruments and manuals should be changed accordingly.

The evaluators' manual should be revised to include an extended discussion on the "quality" of research, since in a large scale operation of the system it may not always be possible to have as many experienced researchers on the panels as would be desirable. This discussion should especially elaborate criteria for judging the quality of evaluation reports submitted by developers in support of the effectiveness of their products. These criteria will be quite similar to those specified for the evaluation of research reports, and should be quite familiar to experienced researchers. It would not hurt to reemphasize these criteria in a separate section specifically discussing the assessment of support documents however. This recommendation may also be problematical, however, in view of the exceedingly low incidence of support documentation obtained in the pilot study.

There should be opportunity for evaluators to make additional written comments on the evaluation booklets. Evaluators requested the opportunity to describe their own professional orientations, i.e., the frames of reference that underlay their judgments. The evaluators felt that the opportunity to do so allowed them more freedom to make certain types of comments regarding products. It was also felt that this type of information on the evaluation form would afford other evaluators better insight into the comments they made.

With regard to modifications of the formal characteristics of the instruments, these deal primarily with production and reproduction considerations. First, all forms and documents used by developers in reporting products and evaluators in evaluating them, were color-coded as to the type of product involved. The primary argument for color-coding is to reduce the possibility of document confusion. There was not a single incidence in the pilot test of this happening. On the other hand, there was some concern over the use of colored forms on the part of laboratory and center respondents. Some felt it would be easier to erase and correct forms if they were on white paper, and others argued that when additional forms were needed, it would be easier to xerox, if white paper were used.

All forms were developed to fit government-size paper, should that be desirable in the future. Laboratory and center respondents were critical of government-size documents, however, because they were not amenable to convenient xeroxing.

Finally with regard to the evaluators' manual, it is strongly recommended that the evaluators' manual be kept in its present size. Its size was designed to make it convenient as a reference handbook during the execution of evaluations. An 8 1/2" x 11" size evaluators' manual would be quite awkward.

In future reproductions of the evaluation manual, however, from cost considerations it is recommended that the manuals be saddle-stitch bound rather than plastic comb bound. This would materially decrease publication costs and greatly facilitate manual storage and shipment via mail.

BUDGETARY RECOMMENDATIONS FOR FUTURE IMPLEMENTATION

Several questions of policy must be considered in determining how to employ the proposed product evaluation system. The answers to these questions will determine to a large extent the mode of utilization of the system.

The first concerns the products to be evaluated. Specifically, should all products be reviewed by an evaluation panel? And if not, what types of products should, or should not, be examined? What criteria should be used to determine whether or not a product should be evaluated?

As previously mentioned in this report, the domain of products ranges from two or three-page journal articles and wall charts to complex and comprehensive educational systems. To carefully evaluate all of these products would require a great amount of time and effort. Unless there is some special purpose to be served, it probably would not be cost effective to treat each product equally. For very inexpensive items, the cost of evaluation could easily exceed the cost to society by letting it go unevaluated. Abbreviated evaluation might be directed to such products, or it might be appropriate, given large quantities of similar items, to assign priorities to the various types of products and only evaluate a subset of them in depth, or to simply sample them, or to prorate the level of evaluation effort according to the level of developmental effort invested.

Further, from a very pragmatic point of view, it might be appropriate to evaluate only those products produced by those programs to which the National Institute of Education is anticipating granting long range funding.

The overriding question to be considered is the amount of resource to be invested in the evaluation of the outcomes of educational research and development. One rule of thumb often cited is that 1 to 2% of the developmental costs of a product should be devoted to its evaluation. Given that nearly $200 million have been spent on laboratories and centers since their inception, this guideline would indicate that $2 million could conceivably be spent on evaluating the products of those agencies. Even half that amount of money is a vast amount. It is probable that no where near that

amount would be reasonable to expect for external evaluation purposes in the near future. In this section the costs of implementing the product evaluation system in a variety of different configurations will be discussed. These configurations have been designed to accommodate variations in the nature of the products to be evaluated and in the composition of the evaluation panel.

Factors affecting costs. The factors most seriously affecting the costs of implementing the evaluation system parallel the policy questions specified above. They are concerned with the products to be evaluated and the composition of the evaluation panels.

With regard to the products, the important considerations are

- the number of products to be evaluated,
- the modes of evaluation to be utilized,
- the amount of travel required for review, and
- the conditions of obtaining the products.

The first, the number of products, is easily understood. Each product to be reviewed increases the amount of time the evaluators must spend, thus increasing the amount of the honoraria to be paid.

The mode in which the product is to be reviewed depends on the nature and availability of the products as well as their complexity. Products which are readily available and self-contained may be examined in the Home/Office Review mode. Products which are extremely complex must be evaluated in the Central Site or Field Review mode. Both the Central Site and the Field Review modes involve greater expenditures than the Home/Office Review in that considerable travel and *per diem* expenses are incurred. In addition, if it is necessary to demonstrate a product, there may be costs of obtaining equipment or bringing the developer to the site where he is to conduct the demonstration. Finally, for products being reviewed sequentially, i.e., that must be passed around, there will be greatly increased costs of communication due to the need for monitoring the progress of the products.

In addition, depending on the availability of the products, it may be necessary to rent or purchase the products, or to pay for shipping them among the different evaluators (as in the case of sequential review).

Regarding the composition of the evaluation panels, there are two factors which should be considered:

- the number of evaluators, and
- their locations.

The influence of the number of evaluators on the costs of implementing the system should require no explanation. The second factor, the location of the evaluators, has several implications for the costs of evaluating the products. First, the costs of travel to and from the training and, perhaps, field visits, will vary depending on the distance the evaluators must travel. Second, if it becomes necessary to negotiate any of the ratings, the costs of either reconvening the panel or holding a conference call will be much greater if the evaluators are far apart. Third, the costs of transporting the products among the evaluators will be increased as the distances between them increase.

In conclusion, then, the costs of the evaluators' honoraria, the evaluators' and the evaluation coordinator's travel and *per diem*, obtaining the products, communications, shipping, and arranging for special equipment and facilities will all vary, depending on decisions made regarding the products to be evaluated and the composition of the evaluation panels.

Bases for computing costs. For the purpose of estimating implementation costs for various configurations of the model, certain assumptions need to be made.

All costs will be expressed in terms of the costs for conducting a single evaluation panel. Certain of the costs will be based on flat rates. Honoraria for evaluators will be $100/day. Honoraria will be given for the day spent in training, for the days spent in reviewing products, and for the day

spent in negotiation. It is assumed that a Home/Office review product can be reviewed in a half-day. A demonstration will be estimated to take one day, and a field trip, two. Negotiation of results is likely to require an additional half-day to day, depending on the number of evaluators and the number of negotiations necessary.

Other costs will be figured based on the experiences gained from the tryouts of the evaluation system. Where travel is necessary and no attempt is made to involve only local personnel, trips will be considered to cost $200 on the average, including ground as well as air transportation. For local travel, the average figure used will be $75/trip. *Per diem*, both for evaluators and the evaluation coordinator's staff, will be $25/day.

Communications, including mailing and shipping as well as telephone charges, will be estimated at $75/month on the average. This figure will be increased if unusually large amounts of communications are necessary. Should conference calls be required, they will be estimated at $70/20-minute call. (This figure assumes 10 stations with a 2,000 mile distance between the farthest ones.) A 20-minute call should suffice for negotiating the ratings of a single product.

Supplies and materials will be estimated at an average figure of $40/month. This includes costs of reproduction services as well as materials. If products must be purchased, an additional average charge of $5.00/product will be assumed. If products must be rented, a fee of $150/day will be assumed.

Finally, the evaluation coordinator's staff time will be charged as follows. Professional staff, including the evaluation coordinator, who would be a senior researcher, and any evaluation associates that are necessary, will be estimated at a rate of $1600/month, or $400/week. Clerical and administrative staff rates will be estimated at $650/month, or $162.50/week.

FIXED COSTS

Costs incurred in selecting evaluators, obtaining products, processing evaluation materials, and preparing reports of the evaluations will not vary with the different configurations of the evaluation paradigm. These costs will be specified first.

Selection of evaluators. The recommended procedure for selecting evaluators is through peer nominations, whereby requests for nominations of people qualified to evaluate products in specific areas are sent to the directors of the R&D agencies, officials of APA and AERA, and other appropriate personnel. The resulting nominations are tabulated and the list of candidates is sent to the directors of the agencies for review. In some instances, such as a lack of consensus in the nominations, a back-up pool of candidates may have to be generated by the evaluation coordinator. Panel members are selected from the list of approved nominees and/or the back-up pool.

An estimated two professional man-weeks would be required to identify appropriate nominators, compile a pool of nominations, review and circulate the list, generate a possible back-up list, select the evaluators, and contact them to confirm their participation. An additional two man-weeks of clerical time would be required to tabulate the nominations as they are received, prepare the lists of nominees for circulation, type the necessary cover letters, and establish a file on each of the selected panel members.

Additional costs incurred would include $100 in charges for supplies, including costs of reproduction of the lists, and substantial communication expenses.

The total estimated cost per panel for the selection of evaluators, then, is $1225.

Obtaining products. Obtaining the products to be evaluated is a two-step procedure. The Evaluation Coordinator first requests a sample copy of

each product for review to determine what evaluation mode will be appropriate for each product. Based on this decision, he will either request a number of copies of the product, or make the necessary arrangements for a demonstration or field visit. He will also ask the developer to update the product report, to clarify any obscure portions of the report, and to send documents supporting statements made about the product. Once the products are received, or the necessary arrangements made, the evaluation coordinator will prepare a master schedule for the conduct of the evaluation and collect and prepare the necessary materials, such as the evaluation forms and instruction manuals.

Approximately one and one-half man-weeks will be required of the evaluation coordinator's time for contacting the developers, reviewing the products, making the necessary arrangements for reviewing the products, and developing the evaluation schedule. An additional one and one-half man-weeks of clerical and administrative support will be necessary for initiating the product log, labeling the elements of the various products, obtaining and labeling the forms, and so forth.

Costs for supplies and communications both will be above average, due to the frequent contacts with the developing agencies and to the need to obtain the evaluation materials.

Total estimated cost for the procurement of ten products for evaluation is $850.

Preparation of evaluation reports. This task involves processing the ratings on the products, computing the necessary statistics, and preparing the evaluation reports. Profiles of the final ratings of each product, as well as summary profiles of all the products, must be developed for inclusion in the reports.

An average of two man-weeks of professional time will be required for processing the data, preparing the profiles, and writing the report. One and

one-half man-weeks, on the average, of clerical support will be necessary for reproducing and circulating the initial ratings, assisting in the development of the profiles, and typing and reproducing the report.

Costs of communications will be average for this task. However, the costs of supplies will be above average due to the necessity of reproducing the evaluation forms and printing the data profiles and the report.

Total estimated cost of data analysis and report preparation is $1125.

Total fixed costs. The following shows the total amount of fixed costs, for a single panel reviewing ten products, for any of the evaluation configurations.

Selection of Evaluators	$1,225
Obtaining Products	850
Preparation of Reports	1,125
TOTAL	$3,200

COSTS OF IMPLEMENTING ALTERNATIVE CONFIGURATIONS

Budgets for five different configurations of the evaluation model will be delineated. For each configuration the specific assumptions on which it is based will be identified, the characteristics of the configuration defined, and the approximate costs of implementation calculated. For the purposes of determining the cost estimates, it will be assumed that each panel will review ten products. One day will be devoted to training in the use of the evaluation materials for each configuration. One-half to one day will be spent negotiating the final ratings.

Standard configuration. In this configuration it will be assumed that the products to be reviewed are typical of the product domain and that no exceptions to the evaluation procedures outlined in this report will occur.

The assumptions underlying this configuration are as follows.

Modes of Evaluation: 1 field visit, 2 demonstrations, and 7 home/office reviews--3 simultaneous and 4 sequential; one demonstration will be conducted by a representative of the developing agency; no special equipment will be required for either demonstration.

Conditions of Obtaining Products: 2 copies of each of 2 products to be purchased; 3 products must be returned, copies of five products supplied *gratis* by the developer.

Required Travel: 2 trips; one to the evaluation coordinator's agency for training and two demonstrations; one for the field visit, plus an additional day to be devoted to any necessary negotiation of final results; a total of 10 days will thus be spent on the road.

Number of Evaluators: 8 members of the core panel; 1 ethnic group representative for home/office review of a product intended for use with minority group children.

Location of Evaluators: it is assumed they come from various sections of the country.

The special tasks of the evaluation coordinator in implementing this configuration are to: plan and conduct the training session; distribute the products for home/office review and follow-up to insure that the products for sequential review are forwarded on schedule; coordinate the demonstrations and the field trip; and return the necessary products at the conclusion of the evaluation effort.

It is estimated that three man-weeks will be required of the coordinator or his staff for these tasks; five days of this time will be devoted to traveling and the field trip, the remainder to the coordination and monitoring of the evaluation activities. Another man-week of clerical support will

be necessary for collecting the necessary materials for the training session, coordinating travel arrangements, processing the evaluation forms, and returning the necessary products. Costs for supplies will be average, but communication expenses will be high due to the extensive monitoring and follow-up of the evaluators, plus the need for returning the three products.

Cost Breakdown:

Evaluation Coordinator's Staff	
Professional -- 3 man-weeks @ $400/week	$1,200
Clerical -- 1 man-week @ $162.50/week	162
Travel -- 1 trip @ $200	200
Per Diem -- 5 days @ $25/day	125
Sub-Total	$1,687
Evaluators	
Honoraria -- 8 for 7.5 days reviewing products, 1 day negotiation, and 1 day training @ $100/day	$7,600
-- 1 for .5 days reviewing products @ $100/day	50
Travel -- 8 for 2 trips @ $200/trip	3,200
Per Diem -- 8 for 10 days @ $25/day	2,000
Sub-Total	$12,850
Supplies and materials (including purchase of 2 copies of 2 products)	30
Communications (including return of 3 products)	65
TOTAL	$14,632
Constant Costs	3,200
GRAND TOTAL	$17,832

Thus the per unit cost for product evaluation would be approximately $1,783. Obviously some economy of scale would accrue with an increase in the number of evaluation panels operating. But even with large scale operation, the unit cost is not likely to be less than $1,500 per item if serious evaluation by a panel of experts is to be realized. The following paragraphs outline a more limited evaluation effort.

Minimal expense configuration. As the title suggests, the objective of this configuration is to minimize the expenses incurred wherever possible without radically deviating from the evaluation paradigm. In order to minimize costs, travel will be curtailed with the exception of one trip for the training session. Field reviews or demonstrations, if necessary, must be conducted at the time of the training session.

The following assumptions prevail.

Modes of Evaluation: 1 field visit, 9 simultaneous home/office reviews;
Conditions of Obtaining Products: agencies will be required to provide sufficient copies of requested products as part of their scopes of work; products will not have to be returned.
Required Travel: 1 local trip will be required for four of the evaluators, and 1 long distance trip will be required for the evaluation coordinator; the evaluators will convene at the site of the field visit where the initial training will be conducted prior to the field review; a total of 5 days will be spent on the road.
Number of Evaluators: 6 evaluators will review each product.
Location of Evaluators: evaluators will be selected to minimize necessary travel costs; thus, all evaluators will be selected from an area relatively near the site of the field visit; as mentioned previously, the evaluation coordinator will travel to that site to conduct the training.

In addition to his tasks of planning and conducting the training sessions and monitoring the evaluators' progress, the evaluation coordinator must arrange for the distribution of the products, either by delivering them in person at the time of training or by mailing them to the evaluators. In addition, because there will be no final field review at which the panel convenes again, the evaluation coordinator is responsible for arranging for negotiation sessions as needed. Only one conference call will be held, during which rating discrepancies for all the products will be discussed.

An estimated two man-weeks will be required of the evaluation coordinator for conducting the training session, coordinating the evaluators' review of the products, and conducting the negotiation session. One week of this time will be spent in traveling for the training session and the field visit. Only one-half man-week of clerical support will be available for processing the forms, arranging for the conference call, and coordinating the travel arrangements.

Communications costs will be well above average with this configuration due to the need for a conference call. Supplies expenses, however, should be somewhat below average.

Cost Breakdown:

Evaluation Coordinator's Staff	
Professional -- 2 man-weeks @ $400/week	$800
Clerical -- 1/2 man-week @ $162.50/week	81
Travel -- 1 trip @ $200	200
Per Diem -- 5 days @ $25/day	125
Sub-Total	$1,206
Evaluators	
Honoraria -- 6 for 6.5 days reviewing products, 1/2 day negotiation, and 1 day training @ $100/day	$4,800
Travel -- 4 local trips @ $75	300
Per Diem -- 4 for 5 days @ $25/day	500
Sub-Total	$5,600
Supplies and Materials	$10
Communications (including 1 7-station conference call estimated at 2 hours)	$250
TOTAL	$7,066
Constant Costs	3,200
GRAND TOTAL	$10,266

Complex products configuration. This configuration explores costs where the maximum number of evaluators is used and where most of the products are sufficiently complex to preclude home/office review. The following are the assumptions relevant to this configuration.

Modes of Evaluation: 4 field visits, 2 demonstrations, 4 simultaneous home/office reviews; some equipment will be required for one of the central site demonstrations; a member of the developing agency's staff will come out to conduct the other demonstration;

Conditions of Obtaining Products: One demonstration product must be rented.

Required Travel: 3 trips; 1 trip to evaluation coordinator's agency for training and two demonstrations; 1 trip to east coast for two field reviews (including one day of travel between sites); 1 trip to west coast for two field reviews (including 1 day of travel between sites); a total of 19 days to be spent traveling.

Number of Evaluators: 9 panel members will review each product.

Location of Evaluators: Evaluators come from various sections of the country.

In this configuration the evaluation coordinator will be less concerned with the mechanics of distributing products and monitoring the evaluators' progress. Most of his attention will be devoted to coordinating the various demonstrations and field visits, as well as conducting the training. Negotiation of ratings can be carried out in conjunction with the various field visits.

Approximately three and one-half man-weeks of the evaluation coordinator's time will be required for coordinating this configuration. He will spend two weeks traveling, for the field reviews. The remainder of the time will be devoted to planning and conducting the training session, arranging and conducting one demonstration, reviewing the results, and conducting the negotiation sessions. Another one and one-half man-weeks of clerical support will be necessary for coordinating the travel arrangements, obtaining the necessary equipment for the demonstration, and processing the evaluation forms.

Costs of supplies will be above average, because of the special equipment that will have to be rented. Costs of communications will be somewhat lower than in the previous configurations, due to the frequent convenings of the panel.

Cost Breakdown:

Evaluation Coordinator's Staff	
Professional -- 3 1/2 man-weeks @ $400/week	$1,400
Clerical -- 1 1/2 man-weeks @ $162.50/week	243
Travel -- 2 trips @ $200	400
Per Diem -- 14 days @ $25/day	350
Sub-Total	$2,393
Evaluators	
Honoraria -- 9 for 12 days reviewing products, 1 day negotiation, and 1 day training @ $100/day	$12,600
Travel -- 9 for 3 trips @ $200	5,400
Per Diem -- 9 for 19 days @ $25/day	4,275
Sub-Total	$22,275
Developing Agency Representative	
Travel -- 1 trip @ $200/trip	$200
Per Diem -- 1 day @ $25/day	25
Sub-Total	$225
Supplies and Materials (including 2 day rental of demonstration product and equipment rental)	$410
Communications	35
TOTAL	$25,338
Constant Costs	3,200
GRAND TOTAL	$28,538

The higher costs of this configuration are more readily acceptable, however, when one considers that this configuration would be used only with more complex products which in turn usually represent very high capital investment in development.

Augmented panel configuration. Several product areas, such as pre-school education programs, are likely to have a relatively large number of products designed for use with, or for the benefit of, minority group students. In such cases, the evaluation panels should include representatives of the appropriate ethnic groups in the consideration of those products. This condition is depicted in this configuration. The panel will be of standard size, and the products typical, as indicated by the following assumptions.

Modes of Evaluation: 1 field visit, 2 demonstrations, 7 home/office review--4 simultaneously and 3 sequentially; no special equipment will be required for the demonstrations; 1 demonstration product, 1 home/office review product, and the field visit will require ethnic group representatives on the panel.

Conditions of Obtaining Products: 2 copies of one product must be purchased; 1 product must be returned.

Required Travel: 2 trips, 1 to the evaluation coordinator's agency for training and the two demonstrations; a second for the field review and any necessary negotiation; in all, 10 days will be spent traveling.

Number of Evaluators: 7 members of the core panel; 2 additional ethnic minority evaluators to review the field review product; another 2 additional evaluators to review 1 demonstration and 1 home/office review product.

Location of Evaluators: evaluators will be drawn from across the country.

The tasks of the evaluation coordinator for this configuration are not appreciably different from his tasks in the standard configuration, with the exception of insuring that the proper representatives are present at the demonstration and field visit. Thus, the costs will differ only to the extent that ethnic group representatives must be accomodated.

Cost Breakdown:

Evaluation Coordinator's Staff	
Professional -- 3 man-weeks @ $400/week	$1,200
Clerical -- 1 man-week @ $162.50/week	162
Travel -- 1 trip @ $200	200
Per Diem -- 5 days @ $25/day	125
Sub-Total	$1,687
Evaluators	
Honoraria -- 7 for 7.5 days reviewing products, 1 day negotiation, and 1 day training @ $100/day	$5,950
-- 2 for 2 days reviewing a product and 1 day training @ $100/day	600
-- 2 for 1.5 days reviewing products and .5 days negotiating	400
Travel -- 7 for 2 trips @ $200/trip	2,800
-- 4 for 1 trip @ $200/trip	800
Per Diem -- 7 for 10 days @ $25/day	1,750
-- 2 for 5 days @ $25/day	250
-- 2 for 3.5 days @ $25/day	175
Sub-Total	$12,725
Supplies and Materials (including the purchase of 2 copies of 1 product)	20
Communications (including the return of 1 product)	50
TOTAL	$14,482
Constant Costs	3,200
GRAND TOTAL	$17,682

Massed review configuration. In this final configuration it will be assumed that all evaluations take place at a central site, such as the evaluation coordinator's agency. (This was the procedure followed during the pilot test of the system.) The emphasis will be on expediting the reviews, in order that the entire task may be completed within a limited time span. Products requiring special consideration, such as a demonstration or field review, will not be evaluated. The following assumptions regarding this configuration describe the nature of the reviews.

Modes of Evaluation: all products will be reviewed in the home/office mode; 2 products will be reviewed during each day, and 1 each night; the remainder of the time will be devoted to training and negotiation; 5 products will be reviewed simultaneously and 5 will be reviewed sequentially.

Conditions of Obtaining Products: 2 copies of 2 products must be purchased; 3 products will have to be returned.

Required Travel: 1 trip to evaluation coordinator's agency where training and review of all products will occur; 7 days will be spent on this trip, including the travel.

Number of Evaluators: 9

Location of Evaluators: 6 evaluators will be drawn from distant states; 3 will come from the local area.

The evaluation coordinator's staff will have a much greater role in this configuration. There will be a far greater need for scheduling, to insure that all the products are reviewed. Similarly, there will be a greater necessity for monitoring the progress of the evaluators, circulating and collecting materials, and supervising the reviews to insure that discussions of ratings do not occur.

Although the review is scheduled to take only one week, it is estimated that two and one-half man-weeks of professional time will be required for supervising and coordinating the reviews. An additional one man-week of clerical support will be necessary for reproducing forms and tabulating results. Costs of communications will be much less for this configuration, as the evaluators will all be present in a central location. Supply costs should be average.

Cost Breakdown:

Evaluation Coordinator's Staff	
Professional -- 2.5 man-weeks @ $400/week	$1,000
Clerical -- 1 man-week @ $162.50/week	162
Travel -- none	
Per Diem -- none	
Sub-Total	$1,162
Evaluators	
Honoraria -- 9 for 3.5 days, .5 days negotiating and 1 day training, @ $100/day	$4,500
Travel -- 6 trips @ $200/trip	1,800
-- 3 trips @ $75/trip	225
Per Diem -- 9 for 5 days @ $25/day	1,125
Sub-Total	$7,650
Supplies and Materials (including purchase of 2 copies of 2 products)	30
Communications (including return of 3 products)	25
TOTAL	$8,867
Constant Costs	3,200
GRAND TOTAL	$12,067

SUMMARY

In order to compare the costs of the various configurations of the evaluation model, the estimated costs of each have been summarized in the following chart. The number of evaluators and amounts of required travel for each configuration have also been indicated, in order to provide a perspective on the cost differences. These figures cover direct costs only. They do not include overhead expenses or fees.

Figure 20

ESTIMATED COSTS OF VARIOUS CONFIGURATIONS

Item	Standard	Minimal Expense	Complex Products	Augmented Panel	Massed Products
Fixed Costs of Preparation	$ 3,200	$ 3,200	$ 3,200	$ 3,200	$ 3,200
Evaluation Coordinator's Expenses--Total	$ 1,688	$ 1,206	$ 2,394	$ 1,688	$ 1,163
Evaluators					
Size of Panel	8(1)*	6	9	7(2)*	9
Honoraria	$ 7,600	$ 4,800	$12,600	$ 6,950	$ 4,500
Number of trips/ Number of people	2/8	1/4	3/9	2/9	1/9
Travel & Per Diem	$ 5,200	$ 800	$ 9,675	$ 5,775	$ 3,150
Total	$12,800	$ 5,600	$22,275	$12,725	$ 7,650
Developer Representative - Total	---	---	$ 225	---	---
Supplies and Materials	$ 30	$ 10	$ 410	$ 20	$ 30
Communications	$ 65	$ 250	$ 35	$ 50	$ 25
Total	$17,783	$10,266	$28,539	$17,683	$12,068

* Figures in parentheses indicate additional evaluator(s) brought in to review one or more specific products.

An examination of the above figures reveals that the expenses of the evaluation coordinator do not vary significantly with the exception of the complex products configuration which requires the evaluation coordinator to participate in two trips rather than one. Similarly, with two exceptions, the cost of supplies and materials and communications are relatively constant. The exceptions occur in the minimal expense configuration when a lengthy conference call is necessary for negotiating final ratings and in the complex products configuration when it is necessary to rent one product as well as special equipment for reviewing another.

The variables most affecting the costs of implementing the various configurations, then, relate to the evaluators. Specifically, the critical variables are the number of evaluators and the amount of travel required. The latter item is, of course, a function of product complexity and the resultant modes

of evaluation. Site visits, field trips and demonstrations not only require more time for review, but also involve considerable expense for travel and *per diem*.

The complex products configuration, requiring a larger number of evaluators, and three trips for each, was by far the most expensive of the configurations. But then it is most apt to be applied only to the most expensive products.

The two configurations involving the least travel, the minimal expense and massed products configurations, were considerably less expensive to implement. The former model also involved a smaller number of evaluators, further reducing expenses. In the latter, because the evaluators conducted all their reviews at a central site, they were requested to review some of the products in the evenings. Thus, more products could be reviewed in a given number of days. This is not reasonable to expect when evaluators are working in their own homes or offices, however.

It is interesting to note that augmenting a core panel with specialists for the review of a specific product or sub-set of products does not affect the overall costs to a large extent. From the above cost estimates the difference between the standard and augmented panel configurations is only $50.

Another point of interest is the differential cost of inviting a representative of the developing agency to conduct a demonstration of a product, rather than sending the panel out to review the product in the field. The costs of bringing in the representative were only $225; to send the panel to the site would cost six to nine times that much.

Factors other than cost should also be considered in determining how to implement the evaluation system. The same procedures which reduce costs may also compromise the quality of the evaluations if adopted uniformly for all product evaluations. For example, imposing tight time constraints on the review of the products, as in the massed products configuration, may result in less thorough examination of the products.

Similarly, configurations involving the use of fewer evaluators, or evaluators selected with regard to geographic convenience to minimize travel expenses, may also result in less valid results for certain products. For example, while brief knowledge products might be quite adequately reviewed by only three or four evaluators, comprehensive educational systems should be carefully and critically considered by a fairly large panel, including specialists representing a variety of areas of expertise.

Thus, the evaluation system submitted herein offers a wide variety of trade-offs between the administrative convenience and the cost of the various configurations and the quality of the resulting product evaluations. These trade-offs, however, can only be weighed, and selected, in the light of government needs as they are defined at a particular point in time.

REFERENCES

The following references were used and/or reviewed in connection with this effort.

Acheson, K. & Olivero, J. Educational laboratories and teacher education. Journal of Teacher Education, 1970, 21, 325-334.

Alkin, M.C. Evaluation theory development. Evaluation Comment, 1969a, 2, (1), 2-7.

Alkin, M.C. Evaluating Cost-Effectiveness of International Programs. Los Angeles: U.C.L.A., Center for the Study of Evaluation, 1969b.

Alkin, M.C. & Woolley, D.C. A model for educational evaluation. Paper presented at PLEDGE Conference, San Demas, California, October 1969.

Allen, J.E., Jr. Strengthening research and development. Remarks at Conference on American Education, Washington, D.C., July 1969.

Alpert, D. The role and structure of interdisciplinary and multidisciplinary research centers. Address to the Ninth Annual Meeting of the Council of Graduate Schools in the U.S., Washington, D.C., December 1969.

Altman, J.W. A scheme for evaluation of the institute for performance technology. (Internal Position Paper.) Pittsburgh: American Institutes for Research, 1966.

Altman, J.W. & Taylor, M.V., Jr. Development of a record form for evaluating research through the report. (Tech. Report #5, ONR Project 153-146) Pittsburgh: American Institutes for Research, 1954.

Altman, J.W. & Taylor, M.V., Jr. Technical appendices for development of a record form for evaluating research through the report. (Tech. Report #5, ONR Project 153-146) Pittsburgh: American Institutes for Research, 1954.

American Institutes for Research. Evaluative Research. Pittsburgh: American Institutes for Research, 1970.

Astin, W. A researcher's proposal for changes in higher education. Educational Record, 1970, 51(3), 225-231.

Bailey, S.K. Significance of the federal investment in educational R&D. Journal of Research and Development in Education, 1969, 2(4), 29-35.

Bendig, A.W. Reliability in the number of rating scale categories. Journal of Applied Psychology, 1954, 38, 38-40.

Bendig, A.W. Rate of reliability in the heterogeneity of scale anchors. Journal of Applied Psychology, 1955, 39, 37-39.

Bloom, B.S. Research and development centers: promise and fulfillment. Journal of Research and Development in Education, 1968, 1(4), 181-189.

Bowman, H.L. The product development process. Paper presented at Regional Council Meeting of Southeastern Education Laboratory, Atlanta, June 1969.

Boyan, N.J. The political realities of educational R&D. Journal of Research and Development in Education, 1969, 2, 3-18.

Boyan, N.J. and Mason, W.S. Perspectives on educational R&D centers. Journal of Research and Development in Education, 1968, 1(4), 190-202.

Bright, R.L. & Gideonse, H.D. Education research and its relation to policy, an analysis based on the experience of the United States. Paper presented to the Committee for Scientific and Technical Personnel, OECD, Washington, D.C., 1967. (ERIC ED 018 866.)

Brown, E.K. More effective educational research through the use of process evaluation techniques. Journal of Research and Development in Education, 1970, 3(4), 59-71.

Brownell, S.M. R&D centers and the schools: a reaction to progress reports. Journal of Research and Development in Education, 1968, 1(4), 173-180.

Caldwell, M.S. An approach to the assessment of educational planning. Educational Technology, 1968, 8, 5-12.

Carrese, L.M. & Baker, C.G. The convergence technique: a method for the planning and programming of research efforts. Management Science, 1967, 13(8), B420-B438.

Case, C.M. Application of PERT to large-scale educational research and evaluation studies. Educational Technology, 1969, 4, 79-83.

Chase, F.S. Educational research and development: promise or mirage? Journal of Research and Development in Education, 1968, 1(4), 3-14.

Chase, F.S. The R&D institutional base: what more is needed? Journal of Research and Development in Education, 1969, 2(4), 19-27.

Chase, F.S. The laboratories: 1970 and beyond. Journal of Research and Development in Education, 1970a, 3(2), 104-120.

Chase, F.S. R&D in the remodeling of education. Phi Delta Kappan, 1970b, 51(6), 299-304.

Coleman, J.S. Participant in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

Cook, D. Management control theory as the context for educational evaluation. Journal of Research and Development in Education, 1970, 3(4), 13-26.

Cooperative Educational Research Lab. A Taxonomy of Programmatic Tasks in an Educational Evaluation Facilitation and Coordination System. Northfield, Illinois: Cooperative Educational Research Lab, Inc., 1969.

Cox, D.R. Criteria for evaluation of reading materials. The Reading Teacher, 1970, 24(2), 140-145.

Crane, H.R. A scientist looks at education. Educational Technology, 1970, 10(1), 22-23.

Cross, K. When will research improve education? Research Reporter, 1967, 2(4). (ERIC: ED 025 206.)

Davidson, W.W. & Gibney, T.C. Evaluation of a modern mathematics program. School Science and Mathematics, 1969, 69(5), 364-366.

Dill, W.R. Varieties of administrative decisions. In Leavitt & Pondy (Eds.), Readings in Managerial Psychology. Chicago: University of Chicago Press, 1964.

Dyer, H.S. Participant in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

Dyer, H.S. Can we measure the performance of educational systems? NASSP Bulletin, 1970, 54, 96-108.

Ebel, R.L. (Ed.) Encyclopedia of Educational Research. (4th ed.) London: The Macmillan Company, 1969.

Education USA, Washington Monitor, October 5, 1970, 29.

Educational Evaluation and Intuition. Educational Forum, 1970, 34(3), 323-332.

Edwards, A.L. On Guttman's scale analysis. Educational and Psychological Measurement, 1948, 8, 313-318.

Edwards, A.L. & Kenney, K.C. A comparison of the Thurstone and Likert techniques of attitude scale construction. In M. Fishbein (Ed.) Readings in Attitude Theory and Measurement. New York: John Wiley & Sons, Inc., 1967.

Etzioni, A. Organizational control structure. In J. March (Ed.), Handbook of Organizations. Chicago: Rand McNally, 1965.

Evans, G.C. Supervising R&D Personnel. New York: Management Association, Inc., 1969.

Farquhar, W.W. & Krumboltz, J.D. A check list for evaluating experimental research in psychology and education. Journal of Educational Research, 1959, 52(9), 353-354.

Federal Research and Development Programs: The Decision-Making Process, House Report No. 1664. Washington, D.C.: U.S. Government Printing Office, 1966.

Feldman, J. & Kanter, H. Organizational decision making. In J. March (Ed.) Handbook of Organizations. Chicago: Rand McNally, 1965.

Festinger, L. & Katz, D. (Eds.) Research Methods in the Behavioral Sciences. New York: Holt, Rinehart & Winston, 1953.

Fiedler, F.E. Engineer the job to fit the manager. Harvard Business Review, September-October 1965, 115-122.

Finn, J.D. Institutionalization of evaluation. Educational Technology, 1969a, 9, 14-23.

Finn, J.D. Evaluation of instructional outcomes: extensions to meet current needs. Version of this paper presented at the annual meeting of the American Educational Research Association, Los Angeles, February 1969b.

Flanagan, J.C. The uses of educational evaluation in the development of programs, courses, instructional materials and equipment, instructional and learning procedures and administrative arrangements. In R.W. Tyler (Ed.) Educational Evaluation: New Roles, New Means. Yearbook of the National Society for the Study of Education, 1969.

Flanagan, J.C. & Jung, S.M. Evaluating a comprehensive educational system. Paper presented at The American Institutes for Research Seminar, Washington, D.C., January 1970.

Freed, M.N. The regional education laboratories: bane or benefit? School and Society, 1967, 95(2298), 494.

Functions of program evaluation and evaluation models in education. High School Journal, 1970, 53(7), 385-399.

Gallup, G. Second annual survey of the public's attitude toward the public schools. Phi Delta Kappan, 1970, 52(2), 97-112.

Geis, G.L. Three innovative teaching systems: common sources of effectiveness. The Record, 1970, 71(3), 379-389.

Gephart, W.J. Profiling educational research. Occasional Paper #3. Bloomington, Ind.: Phi Delta Kappa Research Service Center, 1969a.

Gephart, W.J. The eight general research methodologies: a facet analysis of the research process. Occasional Paper #6. Bloomington, Ind.: Phi Delta Kappa Research Service Center, 1969b.

Gephart, W.J. & Bartos, B.B. Profiling instructional package. Occasional Paper #7. Bloomington, Ind.: Phi Delta Kappa Research Service Center, 1969.

Gideonse, H.D. Research, development, and the improvement of education. Science, 1968, 162, 541-545.

Gideonse, H.D. The OECD policy review of U.S. educational R&D. Educational Researcher, 1970, 21, 5-8.

Ginther, J.R. Can a computer evaluate? School Review, 1971, 79, 603-613.

Glaser, R. Evaluation of instruction and changing educational models. From the proceedings of the Symposium on Problems in the Evaluation of Instruction, U.C.L.A., December 1967.

Glaser, R. Discussion: new myths and old realities. Harvard Educational Review, 1968, 38(4), 739-746.

Glass, G.V. Glass disagrees with Hopkins on means to meet crucial RD&D manpower needs. Phi Delta Kappan, 1969, 51(1), 48.

Guba, E.G. The place of educational research in educational change. Paper presented to the Canadian Council for Research in Education, Winnipeg, Manitoba, June 1967a.

Guba, E.G. Development, Diffusion, and Evaluation. Bloomington, Indiana: National Institute for the Study of Educational Change, October 1967b. (ERIC: ED 015 534.)

Guba, E.G. Evaluation and Change in Education. The National Institute for the Study of Educational Change, 1968.

Guba, E.G. An overview of the evaluation problem. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, February 1969.

Guba, E.G. & Clark, D.L. SEC Newsletter, 1965, Ohio State University, 1(2), 2-5.

Guba, E. & Elam, S. (Eds.) The Training and Nurture of Educational Researchers. Sixth Annual Phi Delta Kappa Symposium on Educational Research. Bloomington, Indiana: Phi Delta Kappa, Inc., 1965.

Guetzkow, H. Communications in organizations. In J. March (Ed.) Handbook of Organizations. Chicago: Rand McNally, 1965.

Guilford, J.P. Psychometric Methods. New York: McGraw-Hill Book Company, Inc., 1954.

Hammond, R.L. Evaluation at the local level. Tucson, Arizona: Project EPIC, 1967 (ERIC: ED 016 547)

Hare, V.C.,Jr. Systems Analysis: A Diagnostic Approach. New York: Harcourt, Brace, & World, Inc., 1967.

Hartley, H.J. Educational Planning--Programming--Budgeting: A Systems Approach. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1968.

Hartup, W.W. Early childhood education and research: significance and needs. Journal of Teacher Education, 1970, 21, 23-33.

Havelock, R.G. et. al. A comparative study of the literature on the dissemination and utilization of scientific knowledge. University of Michigan, Center for Research on Utilization of Scientific Knowledge. (ERIC: ED 029 171)

Hawkridge, D.G. & Chalupsky, A.B. Participants in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

Hawkridge, D.G., Chalupsky, A.B., and Roberts, A.O.H. A study of selected exemplary programs for the education of disadvantaged children, parts I and II. Palo Alto, Calif.: American Institutes for Research, September 1968.

Hemphill, J.K. Educational development. The Urban Review, 1969, 4(2), 23-27.

Hemphill, J. Management and operations of educational laboratories. Journal of Research and Development in Education, 1970, 3(2), 65-80.

Herzberg, F. One more time: how do you motivate employees? Harvard Business Review, 1968, 46(1), 53-62.

Hilgard, E.R. The problem of R&D within behavioral social sciences. Journal of Research and Development in Education, 1969 2(4), 37-48.

Hodges, C.S. Measuring Educational Research Quality and its Correlates. New York: Bureau of Applied Social Research, Columbia University, 1967.

Hoehn, L.P. The regional educational laboratories as change agents. (Doctoral dissertation, Michigan State University) Ann Arbor, Michigan: University Microfilms, 1967. No. 68-7902. (ERIC: ED 035 096)

Howell, R.A. A fresh look at management by objectives. Business Horizons, 1967, 10, 51-58.

Ivancevich, J.M., Donnelly, J.H., & Lyon, H.L. A study of the impact of management by objectives on perceived need satisfaction. Personnel Psychology, 1970, 23, 139-151.

Jay, A. Management and Machiavelli. New York: Holt, Rinehart, and Winston, 1967.

Johnson, G.B.,Jr. A method for evaluating research articles in education. Journal of Educational Research, 1957, 51, 149-151.

Kenefick, R. & Paznak, J. From research to development: education comes of age. The Urban Review, 1970, 4(3), 24-28.

Kohr, R.L. & Suydam, M.N. An instrument for evaluating survey research. The Journal of Educational Research, 1970, 64(2), 78-85.

Komoski, P. The continuing confusion about technology and education. Educational Technology, 1969, 9(11), 70-74.

Kresh, E. An overview of the discrepancy evaluation model and a related case study. In J. Davis & M. Essex (Eds.) Educational Evaluation. Conference proceedings, Ohio Department of Education, July 1969.

Kresh, E. Evaluation in a multi-level agency. Journal of Research and Development in Education, 1970, 3(4), 97-102.

Light, R.J. & Smith, P.V. Choosing a future: strategies for designing and evaluating new programs. Harvard Educational Review, 1970, 40, 1-28.

Lindblom, C.E. The science of muddling through. In Leavitt and Pondy (Eds.), Readings in Managerial Psychology. Chicago: University of Chicago Press, 1964.

Lundin, E.G. & Welty, G.A. The relevance of a managerial decision to educational administration. Paper presented at the Annual Meeting of the American Educational Research Association, Minneapolis, March 1970a.

Lundin, G.E. & Welty, G.A. Management models and educational evaluation. Journal of Research and Development in Education, 1970b, 3(4), 38-47.

Madden, J.M. & Bourdon, R.D. Effects of variations of rating scale format on judgment. Journal of Applied Psychology, 1964, 48, 147-151.

Mann, J. Participant in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

March, J.G. Business decision-making. In Leavitt and Pondy (Eds.) Readings in Managerial Psychology. Chicago: University of Chicago Press, 1964.

Marrow, Bowers & Seashore. Management by Participation. New York: Harper and Row, 1967.

Marshall, M.S. Analyzing evaluation. Educational Forum, 1971, 35, 487-490.

Massie, J.L. Management theory. In J. March (Ed.) Handbook of Organizations. Chicago: Rand McNally, 1965.

Mayer, M. Participant in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

McIntyre, R.B. & Nelson, C.C. Empirical evaluation of instructional materials. Educational Technology, 1969, 9(2), 24-27.

McNemar, Q. Opinion-attitude methodology. Psychological Bulletin, 1946, 43(4), 289-374.

Merriman, H.O. From evaluation theory into practice. Journal of Research and Development in Education, 1970, 3(4), 48-58.

Merwin, J.A. Evaluation designs and instruments. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, February 1969.

Metfessel, N.S., and Michael, W.B. A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. Educational & Psychological Measurement, 1967, 27, 931-943.

Morgenstern, D., Shephard, R.W., & Grabowski, H. A graph oriented model for research management. In M.C. Yovits, et. al. (Eds.) Research Program Effectiveness. New York: Gordon & Breach, Science Publishers, Inc., 1966.

Myers, A.E. The impact of evaluative research on educational programs for the poor. Teachers College Record, 1970, 71, 371-378.

National Center for Educational Research and Development. Educational Research and Development in the United States, Washington, DC: USOE Final Report, OEDC Policy Review, December 1969.

Nicodemus, R.B., et. al. The 'Pygmalion effect' in science curriculum implementation. Rockville, Maryland: Montgomery County Public Schools, 1969. (ERIC: ED 933 867).

Odiorne, G.S. Management by Objectives. New York: Pitman Publishing, 1965.

Parker, L.A. The structure, management, and development of federally-funded "permanent" educational research and development organizations. (Doctoral Dissertation, Revised Version, Graduate School of Education, Harvard University) Cambridge, Mass., 1970.

Phi Delta Kappa National Study Commission on Evaluation. Educational Evaluation and Decision-Making, 1970.

Pinney, R.H., et. al, Research and development: the reduction of uncertainty. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, February 1969. (ERIC: ED 028 121).

Popham, W.J. Focus on outcomes--a guiding theme of ES '70 schools. Phi Delta Kappan, 1969, 51, 208-210.

Provus, M. Evaluation of ongoing programs in the public school system. In R.W. Tyler (Ed.) Educational Evaluation: New Roles, New Means. Yearbook of the National Society for the Study of Education, 1969.

Provus, M. Evaluation for decision-making, choice, and value. Journal of Research and Development in Education, 1970, 3(44), 1-12.

Raia, A.P. A second look at management, goals, and controls. California Management Review, 1966, 8, 49-58.

Raisbeck, G., et. al. Management Factors Affecting Research and Exploratory Development. Cambridge, Massachusetts: Arthur D. Little, Inc., 1965.

Randall, R.S. An operational application of the Stufflebeam-Guba CIPP model for evaluation. Paper read at the Annual Meeting of the American Educational Research Association, Los Angeles, February 1969a.

Randall, R.S. Knowledge about decision processes and information. Paper read at the Annual Meeting of the American Educational Research Association, Los Angeles, February 1969b.

Reagan, M.D. Basic and applied research: a meaningful distinction? Science, 1967, 155, 1383-1386.

Research and development: advances in education. U.S. Office of Education research document, 1968.

Rhodes, L. Linkage strategies for change: process may be the product. Phi Delta Kappan, 1969, 51, 204-207.

Rice, D., Buser, R., & Ellis, J. Educational evaluation--a model for task-oriented position development. Contemporary Education, 1970, 41(3), 115-118.

Riker, B.L. A comparison of methods used in attitude research. Journal of Abnormal and Social Psychology, 1944, 39, 24-42.

Rossi, P. Participant in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

Rozeboom, W.W. & Jones, L.V. The validity of the successive intervals method of psychometric scaling. Psychometrika, 1956, 21(2), 165-183.

Satlow, I.D. Do's and don't's of evaluation. The Journal of Business Education, 1967, 42, 274-276.

Schutz, R.E. The nature of educational development. Journal of Research and Development in Education, 1970, 3(2), 39-64.

Scriven, M. Participant in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

Shalock, H.D. Research and evaluation: data generating activities that vary in purpose, design, and output. Paper presented at the Annual Meeting of the American Educational Research Association, Minneapolis, March 1970.

Shaw, M.E. & Wright, J.M. Scales for the Measurement of Attitudes. New York: McGraw-Hill Book Company, 1967.

Shaycoft, M.F. & Altman, J.W. A procedure for evaluating graduate research on the basis of the thesis. (Tech. Report #7, ONR Project 153-146), Pittsburgh, PA: American Institutes for Research, 1955.

Shepard, H.A. Patterns of organization for applied research and development. In Leavitt and Pondy (Eds.) Readings in Managerial Psychology. Chicago: University of Chicago Press, 1964.

Sieber, S.D. Organizational resistances to innovative roles in educational organizations. Columbia University, Bureau of Applied Social Research. Paper prepared for UCEA Career Development Seminar, Portland, October 1967. (ERIC: ED 015 536)

Sieber, S.D. & Lazarsfeld, P.F. The Organization of Educational Research, Columbia University, Bureau of Applied Social Research, 1966.

Sjogren, D.D. Measurement techniques in evaluation. Review of Educational Research, 1970, 40(2), 181-212.

Sorenson, G. A new role in education: the evaluator. Evaluation Comment, 1968, 1(1), 1-4.

Stake, R.E. The countenance of educational evaluation. Teachers College Record, 1967, 68(7), 523-540.

Stake, R.E. Objectives, priorities, and other judgment data. Review of Educational Research, 1970, 40(2), 181-212.

Stake, R.E. & Denny, T. Needed concepts and techniques for utilizing more fully the potential of evaluation. In R.W. Tyler (Ed.) Educational Evaluation: New Roles, New Means. Yearbook of the National Society for the Study of Education, 1969.

Strauss, S. Guidelines for analysis of research reports. The Journal of Educational Research, 1969, 63(4), 165-169.

Stufflebeam, D.L. Toward a science of educational evaluation. Educational Technology, July 1968, 5-12.

Stufflebeam, D.L. An emergent theory of evaluation. Paper read at the Annual Meeting of the American Educational Research Association, Minneapolis, March 1970.

Suchman, E.A. The logic of scale construction. Educational and Psychological Measurement, 1950, 10, 79-93.

Suchman, E.A. Participant in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

Symonds, P.M. A research checklist in educational psychology. The Journal of Educational Psychology, 1956, 47(1), 100-109.

Torgerson, W.S. Theory and Methods of Scaling. New York: John Wiley & Sons, Inc., 1958.

Townsend, R. Up the Organization. New York: Alfred A. Knopf, 1970.

Trow, M. Methodological Problems in the Evaluation of Innovation. Los Angeles: U.C.L.A., Center for the Study of Evaluation, 1969.

Umans, S. The Management of Education. Garden City, New Jersey: Doubleday & Company, Inc., 1970.

United States Office of Education. Educational Research and Development in the United States. Document number OE-12049, December, 1969.

United States Office of Education. U.S. Office of Education 1970 Evaluation Plan. Unpublished manuscript. Washington, DC: Office of the Deputy Assistant Secretary for Planning, Research, and Evaluation, 1970.

Wandt, E., et al. An evaluation of educational research published in journals. Report of the Committee on Evaluation of Research. Washington, DC: American Educational Research Association, 1967.

Welty, G.A. Research design and evaluation research. Paper read at the Annual Meeting of the American Educational Research Association, Los Angeles, February 1969a.

Welty, G.A. The Logic of Evaluation, Chatham College; Educational Resources Institute, October 1969b.

Westbury, I. Curriculum evaluation. Review of Educational Research, 1970, 40(2), 239-260.

Whatever happened to those regional laboratories? Grade Teacher, 1969, 87(4), 96-100.

White, M.H. VALED: A system for evaluative educational decisions. Paper from which earlier version was presented at the Annual Meeting of the American Educational Research Association, Minneapolis, March 1970.

Wiley, D.E. The Design and Analysis of Evaluation Studies. Los Angeles: U.C.L.A., Center for the Study of Evaluation, 1969.

Wrightstone, J.S. Participant in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

Wynne, E. Participant in "Evaluating Educational Programs: A Symposium." The Urban Review, 1969, 3(4).

APPENDICES

Appendix A

INSTRUCTIONS FOR COMPLETION OF PRODUCT REPORTING FORMS

AND

SAMPLE PRODUCT REPORTING FORMS

Instructions For Completion Of

Product Reporting Forms -

Developmental Products

Form 10-71-A

Preface

AN OVERVIEW OF THE NEW NCERD PRODUCTS REPORTING SYSTEM

For a number of months the Division of Research and Development Resources has been working toward the development of a unified, comprehensive products reporting system which would adequately reflect the broad spectrum of Laboratory and Center accomplishments. The position of Products Coordinator was established at Division staff level and planning was begun.

Three broad categories of Center and Laboratory outcomes have been defined. The first accommodates those products which have typically been described as "hard" products, i.e., products deriving from systematic developmental efforts and which often (although not necessarily) have some commercial value. Examples of such *developmental* products are: curricular materials; work books; teacher training programs; career games; toy libraries; etc.

The second category of Laboratory and Center outcomes encompasses those efforts at the production of *new knowledge*, i.e., at the expansion of the knowledge base on which new educational development efforts might be based. Knowledge products may take the form of: research reports; reviews of literature never before summarized; new theoretical models; evaluation studies; the creation of new conceptual systems; and the like. The crucial factor here is that it is either "new" knowledge, or old knowledge synthesized in a form not hitherto available.

The third category of outcomes deals with those Laboratory and Center outcomes concerned with improving what might be called the "state of the educational R&D art," i.e., institutional capability for R&D in the United States. Products of this type may be much less tangible than those above, but not necessarily less valuable. Such *institutional capability* products might include: an increased R&D manpower base, through staff development and researcher training; the development of cooperative research, communication, and dissemination networks; the development of educational R&D management expertise; catalytic effects, through visible leadership on educational R&D activities; and the like.

Different forms will be used for reporting contributions (products) in each of these areas. The system has been entitled the PARaDE (Products/Accomplishments from Research and Development in Education) system. It is currently planned to have annual reporting with periodic updating as warranted by product development.

The purpose of the system is to provide NCERD with a single, authoritative source of information about <u>all</u> Laboratory and Center products.

It is expected that PARaDE information will be especially useful to NCERD, NCEC, OPE, and other governmental agencies, as an initial source of information about Laboratory and Center products. While the system clearly will not provide all information that any potential user might eventually need, it is felt the PARaDE system will materially reduce the number of product information requests that Laboratories and Centers receive from various governmental agencies and contractors and will minimize the number of conflicting reports often heard regarding Laboratory and Center accomplishments which result from differences in data sources, product definitions, reporting procedures, and the like.

General Information

- WHAT IS A DEVELOPMENTAL PRODUCT?

A product is defined as a solution to an educational problem. A developmental product, then, provides materials or other "goods" which are needed in education. Developmental products are goods which can be marketed (in the most general sense) and disseminated to schools and/or other consumers. These products often take the form of text books, films, manuals, tests, tapes, and other instructional materials. Research findings and evaluation studies, though published in the form of diffusible reports, do not constitute developmental products as they are developed in response to a need for new knowledge rather than new goods.

- WHAT IS AN EDUCATIONAL PROBLEM?

An educational problem is defined as a need for a product which will accomplish a specific goal. An example of a problem is "Teacher educators need materials which will permit the individualized training of teacher-trainees in the use of reinforcement techniques in the classroom."

You should be careful not to define the various problems your agency is addressing too broadly or too narrowly. An example of a problem that is too broadly defined is "There is a need for individualized educational programs." This statement embodies a whole complex of problems, such as a need for curricular materials that can be organized and structured for use in individualized programs, a need for training programs to train teachers to individualize their instruction, and so forth. It is better to conceptualize such comprehensive areas in terms of their functional components.

You should also take care to avoid the other extreme where problems are defined at such low levels that they appear, for all practical purposes, as trivial. For instance, "a need for a student answer key for the XYZ Achievement Test" and "a need for a student workbook to go with a 10th grade social studies text" do not reflect very significant problems.

In identifying the problems your agency is addressing, then, you should define them at a moderate level of specificity, neither too narrowly nor too broadly. Problems should be defined narrowly enough to be manageable, yet still broadly enough to be meaningful.

- WHY ARE THERE LIMITS ON THE "ACCEPTABLE" RANGE OF SPECIFICITY OF PROBLEMS?

The primary reason for limiting the range of definition of products is to facilitate product reporting. If you define problems too narrowly, you may end up reporting on every item produced. On the other hand, if you define problems too broadly, an inordinate number of man-years of effort may be spent without any apparent output. By defining problems at a moderate level of specificity you will be able to report a reasonable number of products which could still be judged significant.

- HOW DO I DETERMINE WHAT MY DEVELOPMENTAL PRODUCTS ARE?

Developmental products should be defined at the lowest possible level at which they represent complete functioning units. That is, a "product" should include all the elements necessary for its use or operation. Thus, you would not want to consider each manual, workbook, and test booklet for a reading program as separate products since they all function together in the operation of the program. On the other hand, if you are developing an instructional "system" you should consider the reading program, mathematics program, science program, staff development program, and so forth, as separate products since they could all function independently of the others, though together they comprise an elementary-level educational system. In identifying your developmental products, then, you should select those that constitute single but complete units. (This discussion is elaborated more fully in the instructions for SECTION 4.)

- HOW MANY PRODUCTS SHOULD I REPORT?

In reporting your agency's products you should complete a form on each product that has been completed or is presently under development. For the purpose of this form "completed" refers to the conclusion of your agency's responsibility in developing the product, even if this only involves preparing a prototype for field-testing. The scope of an agency's responsibility for its products may vary from product to product. In some cases an agency may be responsible for the entire development of a product, from developing a prototype to preparing it for marketing and dissemination; in other cases one agency may enter into a cooperative relationship with another agency, whereby one is responsible for initial developmental efforts and the other is concerned with the tryout and revision of the product; in still other cases an agency may develop a prototype product and then revise the product after another agency has field tested it. In reporting your agency's products, then, you should be sure to include any products which your agency is involved in, not just those for which your agency is entirely responsible. You should not report products which you are planning on developing but have not as yet begun. Note: a separate form should be filled out for each product.

- HOW WILL ALL THE WORKBOOKS, TESTS, ETC. THAT I DEVELOP SHOW UP IF I DON'T COMPLETE FORMS ON THEM?

The various pieces comprising a product, such as instructions manuals, booklets, computer programs, and so forth should be listed in Section 13 - Product Elements on the form. They will be considered as elements of the overall product.

- CAN I SEND PARAGRAPHS AND PAGES FROM OTHER OF OUR DOCUMENTS?

It would generally be quite unwise to do so. As you complete this form, you will be asked to observe fairly specific instructions related to each question. Abbreviated examples will illustrate these instructions. It is highly unlikely that "cutting and pasting" from pamphlets, brochures, annual reports, etc., would respond specifically to the instructions.

If however, you feel that a particular document provides additional support or elaboration for the material you have written, you are encouraged to cite it as a support document. Support documents, i.e., documents providing additional support for, or explanation of, your product, are especially desirable in the problem, strategy, outcomes, potential consequences, market, and product description. When you cite support documents be sure to identify them completely.

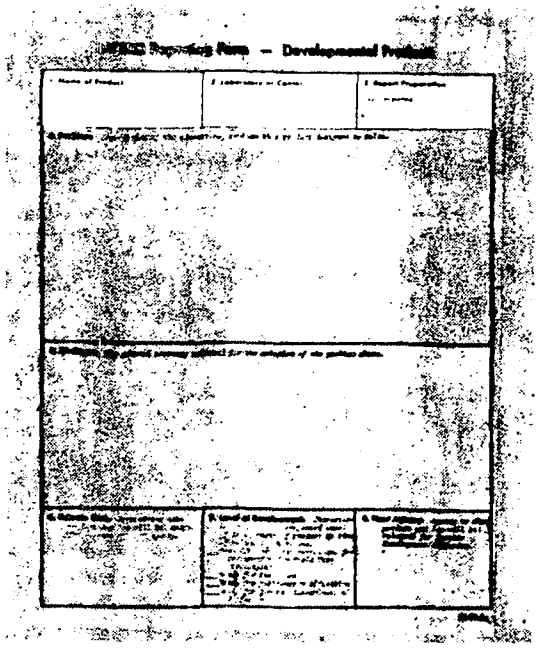
- WHAT IF I NEED MORE SPACE TO ANSWER THE QUESTIONS?

If you need more space than is provided on the form, continue your answer on a separate page, which should then be attached to the form. Be sure to indicate on the form that you are continuing your response on the appended pages by writing "continued on attached page."

- - - - - - -

If you have any questions concerning the completion of this form, please call the Product Coordinator at (415) 328-3530, ext. 900.

SECTION 1. Name of Product
SECTION 2. Laboratory or Center
SECTION 3. Report Preparation



Instructions

SECTION 1. NAME OF PRODUCT

List the name of the product and any acronym used in referring to it.

SECTION 2. LABORATORY OR CENTER

Write the full name of your Laboratory or Center in this section, giving the acronym as well. In subsequent discussions the word "agency" will be used as a synonym for both Laboratory and Center.

SECTION 3. REPORT PREPARATION

Indicate the date on which this form was completed. After the form has been filled out, it should be reviewed by the agency Director and signed by him, or his delegate, before it is transmitted to OE.

Clear Examples	Less Clear Examples
SECTION 1. NAME OF PRODUCT	
1.1 Research Utilizing Problem Solving (RUPS)	1.2 RUPS
2.1 Minicourse 1: Effective Questioning, Elementary Level	2.2 Minicourse 1
SECTION 2. LABORATORY OR CENTER	
1.1 The Center for the Advanced Study of Educational Administration (CASEA) at the University of Oregon	1.2 The Oregon Center
2.1 Southwest Educational Development Laboratory (SEDL)	2.2 The Southwest Laboratory

3.1 The Research and Development Center for Teacher Education (Stanford University)	3.2 Teacher Ed. Research Center

SECTION 3. REPORT PREPARATION

1.1 Date prepared September 15, 1971. Reviewed by John Smith, Agency Director.	1.2 Date prepared Sept 15 Reviewed by director
2.1 Reviewed by John Doe for John Smith, Agency Director.	2.2 Reviewed by JD for JS
3.1 Reviewed by Jack Brown, Dissemination Officer.	3.2 Reviewed by the dissemination office

Comments

SECTION 1. NAME OF PRODUCT

The less clear examples are simply abbreviations of the complete titles. It will be helpful to supply an acronym for quick reference if the product is later referred to under that designation.

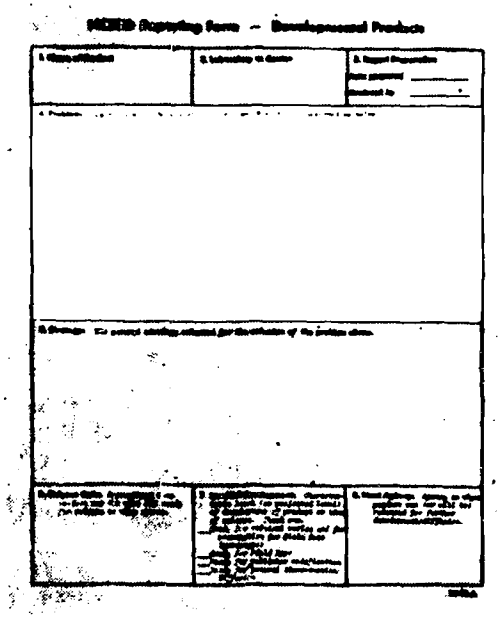
SECTION 2. LABORATORY OR CENTER

This section is quite self-evident. Because some agencies have similar names, please give complete names.

SECTION 3. REPORT PREPARATION

This section, also, requires little comment. Please indicate the date, including the year, that this product report was prepared. The second line is for the signature of the agency Director. It is highly desirable for the Director personally to review the description of each product reported by his agency. He may, however, delegate this responsibility if he so chooses. If this should be the case, the individual so designated should sign, giving his full name.

SECTION 4. Problem



Instructions

Describe the educational problem this product is designed to solve. A problem is defined as a recognized discrepancy between an existing state and a desired end state. The problem should be stated at the lowest level which is still meaningful. That is, the lowest level at which the product can stand as a functional unit. For example, while it would be theoretically possible to design a complete reading program for a single grade only, it probably would not be reasonable to do so in that the learning of basic reading skills typically requires several years.

To more precisely define the scope of the problem, indicate the special characteristics of the target group for which the product is intended. For example, "a reading program is needed for primary grade Chicano children."

The problem statement may also specify what is needed to solve the problem. For example, "a multimedia package of materials is needed to teach improved reading skills to Chicano children."

Clear Examples

1.1 There is a need for effective curriculum materials to be used in training teachers to individualize reading instruction in the elementary grades.

2.1 Training materials are needed to train teachers to individualize math instruction in the elementary grades.

3.1 Assessment materials and procedures are needed

1) to determine specific skill areas in which preschool teachers need training and

2) to monitor teacher trainees' progress in an individualized program of studies.

Less Clear Examples

1.2 There is a need for a comprehensive preservice teacher training program.

2.2 A videotape is needed to train teachers to individualize instruction.

3.2 Scoring keys are needed to score interim and final trainee evaluations in the program.

4.1 Instructional materials are needed to train teachers in an individualized computer-managed educational program to use a computerized data management system for keeping student records.	4.2 Teachers don't keep records well.
5.1 Teachers who are being prepared to work in schools with large numbers of black students need training in black history and culture, black language patterns, and in skills for relating to the black community.	5.2 Teachers who are being prepared to work in inner-city schools need special training.
6.1 Inquiry training materials and techniques are needed to prepare teachers to utilize process oriented science and social studies curriculum materials. (See "New Directions in Teacher Training," Robert K. Brown, Journal of Teacher Training, 2, 1969, p66-75, for a fuller statement of problem being addressed by this product.)	6.2 Teachers are not prepared in inquiry training.

Discussion of Examples

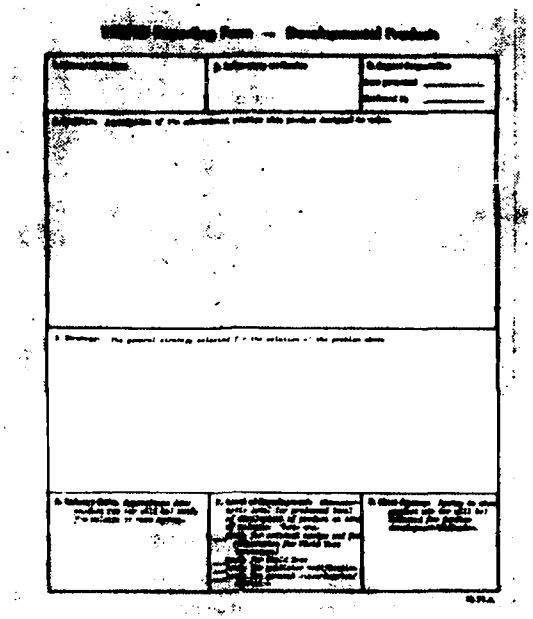
All of the examples above relate to aspects of a comprehensive preservice teacher training program. However, you shouldn't state the problem as a need for a comprehensive teacher training program (1.2); such a statement defines the problem too broadly. As you can see in examples 1.1 to 4.1, a variety of problems can be identified in this area. On the other hand, you should avoid stating the problem too narrowly, as in examples 2.2 and 3.2. The videotape is only one part of the package of training materials. Similarly, a test scoring key is not a functionally independent unit; it only becomes useful in conjunction with the test, answer sheets, and other assessment materials.

You should also avoid vagueness in stating the problem. Examples 4.2 to 6.2 may indeed be problems, but give no indication of what can be done. They simply report an understandable state of affairs. Your problem statements should, then, be phrased positively, suggesting what can be done, not simply what is wrong; that is, they should state an explicit need for <u>something</u> to solve a problem.

By way of reiteration, be careful not to define the problem either too narrowly or too broadly. If the product solution cannot stand alone as a meaningful functional unit, then you have defined the problem too narrowly. Examine the problem statement carefully and determine what additional aspects need to be included to provide for a discrete and independent product solution.

Conversely, if the product solution consists of a number of functionally independent units, the problem has been defined too broadly. In this case you should examine your problem statement and identify the component problems having functionally independent product solutions. Each of these component problems should be listed on separate product reporting forms. Remember, the problem should be stated at the lowest possible, yet still <u>meaningful</u>, level of specificity.

SECTION 5. Strategy



Instructions

This section should state what overall strategy is being taken in the solution of the problem. You should describe as explicitly as possible the approach being taken and why this approach was selected over others. Your discussion should present the operating guidelines followed in developing the product. If other agencies are to be involved in the product's development, their involvement should be explained here. You may find it helpful to briefly mention what alternative strategies were considered; however, this information is not required.

You should note that this section is not an overview of the product itself -- its characteristics, intended outcomes, and so forth are described later -- but rather a description of the general approach, or operating and administrative guidelines, you followed in developing the product.

Clear Examples

1.1 The training materials are to be modular in nature (to allow for individualization of content), criterion referenced (to allow trainees to challenge a module if they already know the material it covers) and performance based (to enable trainers and trainees to determine if the module objectives have been met).

Because of our agency's background in large-scale field testing and commercial development of products we decided to enlist the aid of the XYZ agency to do the requisite basic research and to develop prototype materials. We are responsible for testing and refining the materials and for preparing them for dissemination.

Alternative strategies considered included conducting a series of lectures and seminars for the teacher trainees and training teams of inservice-consultants to work with individual teachers on developing techniques for individualizing reading instruction. Some consideration was also given to having our agency responsible for the entire developmental effort.

Less Clear Examples

1.2 Eight modules containing approximately 5,000 work sheets, and pre- and post-criterion referenced tests, are to be developed. Each module will contain five learning episodes; these episodes will contain activities designed to assist the trainees meet the performance objectives of the module. Other agencies will assist in the development of the materials.

2.1 The teacher training materials should model the format and approach of the student materials for two reasons: 1) teachers tend to teach as they are or were taught, so they should be taught with individualized materials; 2) individualization of teacher training materials allows for variations in content and pace of instruction for individual teachers. It is expected that we will enter into a cooperative relationship with the ABC agency, whereby we will design and develop the prototype materials and they will prepare them for field testing and will revise the materials.

3.1 To produce a set of training materials which will be packaged so as to minimize the amounts of time required of trainees and still provide a variety of materials. Providing varied materials has been found to produce greater interest in the materials and to result in more efficient learning. Our agency will be responsible both for the initial development of the materials and for testing them and preparing them for dissemination. (For elaboration of strategy underlying this product, see Technical Memo 84. "Procedures for the Development of Teacher-Training Materials," by John D. Jones, 1969.)

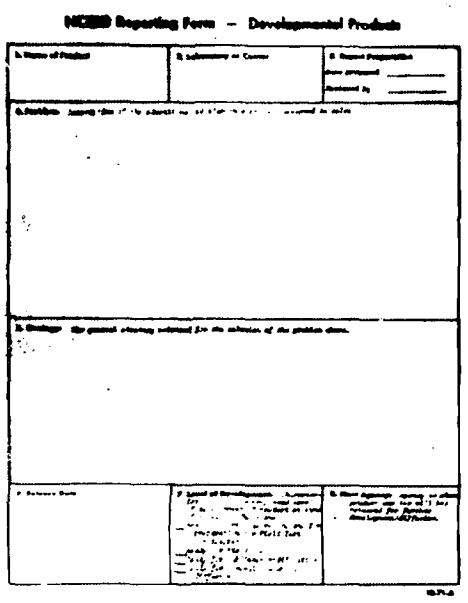
2.2 We will develop teacher training materials. These materials should be individualized.

3.2 To produce materials to train teachers to implement an individualized reading program.

Discussion of Examples

All of the examples above relate to strategies for solving the problem of a need for training materials to teach teachers how to individualize reading instruction in the elementary grades. Examples 1.1, 2.1, and 3.1 state specifically what strategy is to be followed in developing the product (the teacher training materials). Example 1.2, on the other hand, makes vague references to the strategy, but it basically is a description of the product whereas example 1.1 makes the strategy explicit and avoids detailed descriptions of the product itself. Moreover, example 1.2 simply mentions that another agency will be involved while example 1.1 specifies both its own domain of responsibility and that of the other agency. Finally, example 1.1 also mentions other strategies that were considered for providing solutions to the problem.

Examples 2.2 and 3.2 also do not provide clear statements of the strategy to be adopted. Example 2.2 mentions only that the agency will develop materials and that these materials should be individualized; a complete statement of the strategy should include reasons why the materials are to be individualized and should have specified the scope of the agency's responsibility for the development of the product, as in example 2.1. Example 3.2 simply reiterates the general purpose of the product; it does not tell what approach was taken in producing the materials.



SECTION 6. Release Date

Instructions

In this section you should give the date on which the product was ready, or will be ready, for release. If your agency has completed its effort on the product and has released it to another agency, such as another Laboratory or a publisher, or to the actual product users, you should write the date on which your agency's effort was completed. Note: this does not mean the product must be completely "finished" or ready for broad-scale dissemination or utilization; rather, it only means that your agency's effort in developing the product has been concluded. If your product is ready for release, but has not yet been released, you should indicate the date on which it was ready for release in this section.

If your agency is still engaged in developing the product you should give the date on which you anticipate your agency's effort will be concluded -- the date on which the product will be released to another agency or to consumers.

In providing the date of release or anticipated release, you should be as precise as possible. Whenever possible, you should include at least the month and year of release.

Report a product as released if it has been released for wide-spread adoption, even though you expect to revise and refine the product at some later date.

Clear Examples

1.1 April 15, 1971

2.1 March, 1970

3.1 On or about January 15, 1972

4.1 Sometime between now and January 1, 1972.

5.1 Sometime during the first quarter of 1972 depending on contract negotiations.

Less Clear Examples

1.2 Last year sometime

2.2 Released for internal use only.

3.2 This product is just about ready for release.

4.2 Release pending

5.2 We anticipate product release in the very near future.

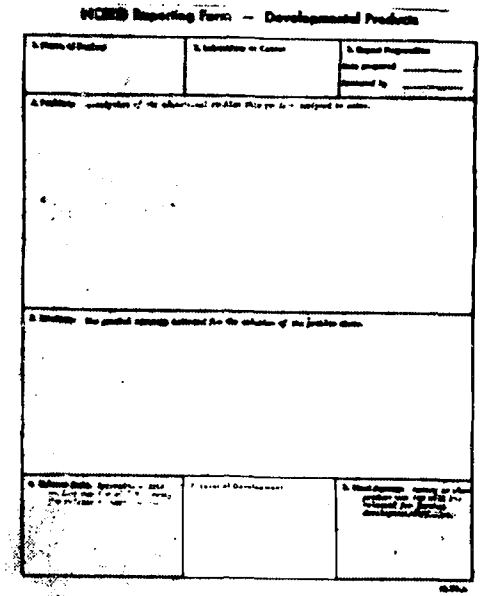
6.1 Model 1 will be released for publication as soon as evaluation is complete - approximately December 1, 1971. However, we expect to systematically revise this product over the next five years.

6.2 Just as soon as the evaluation results are in and analyzed, but we expect to revise these materials continuously over the next five years on the basis of user feedback.

Discussion of Examples

As you can see from the "Clear Examples," the release date may be given in a variety of ways. For products that are already released you should state at least the month and year and the actual date if possible as in examples 1.1 and 2.1. Example 1.2 does not itself provide information about when the product was released; this can only be inferred by noting the date on which the form was completed. Example 2.2 does not really describe a "release" as the agency is still working on the product.

Examples 3.1, 4.1, and 5.1 provide information about when you expect the product to be released. Note that in all cases the year was given to provide a reference point. Examples 3.2, 4.., and 5.2 do not give a precise estimate of when the product will be released; they only assert that the product will be released.



SECTION 7. Level of Development

Instructions

This section should state how far along the product was, or will be, upon release. To fill in this section you should place a mark beside the level of development that most accurately describes your product as it was, or will be, at the time of its release from your agency. You should mark only one level of development for your product; if different aspects of the product were or will be in different stages of development upon completion of your responsibility for the product, you should select the one stage that best characterizes the product as a whole. Please select one of the four categories given on the form; do not write in your own category.

Examples

1. You have prepared a set of social studies modules for further development by another R&D agency. Your work involved identifying reference materials, writing learning activities, and developing assessment exercises. You submitted drafts of your materials to the agency responsible for their furthur development.	You should have checked "ready for critical review and for preparation for field test (ie, prototype materials).
2. You have written a manual instructing teachers in methods of using reinforcement techniques in their teaching. You have had a number of psychologists review the module and you have tried it out with a number of teachers. You have incorporated all the comments into a revised version of the manual, and submitted this version to another agency, which will conduct additional tests with other groups of teachers.	You should have checked "ready for field test."
3. You have prepared a set of videotapes in which teachers model the use of non-verbal communication techniques in the classroom. You have also developed supporting workbooks containing activities by which teachers will practice using these techniques. The entire package has been tested and revised, and you have turned it over to a publisher. The publisher will determine how the materials should be packaged for marketing--whether to use videotape or 16mm film, whether the printed materials should be bound or looseleaf, etc.	You should have checked "ready for publisher modification."
4. In developing a set of mathematics textbooks, you have drafted and tested the books, and have prepared revised versions based on your tests. You have submitted them to a publisher who has suggested certain	You should have checked "ready for general dissemination/diffusion."

modifications in the artwork and suggested some content revisions in one chapter. You have completed these modifications and returned the manuscripts to the publisher for printing and distribution.

5. In preparing a handbook which will introduce new personnel in educational R&D agencies to the kinds of skills that will be required, you have surveyed a number of agencies and prepared a manuscript containing the results of your survey. Your manuscript is in "rough draft" form when you submit it to the R&D agency which requested it--it contains some handwritten corrections and a page of inserts has been appended to the manuscript. The R&D agency will prepare a clearer and usable version of your document for use with their new personnel.

   You should have checked "ready for critical review and for preparation for field test (prototype)."

6. You have produced a series of student activity books designed to introduce the students to various leisure time activities. These books have been tried out with a number of students and some modifications have been made in the original versions. You released the workbooks to a publisher who reviewed the materials and made some additional changes. The workbooks have been printed and are now in use in schools in your state.

   You should have checked "ready for publisher modification."

Discussion of Examples

The following definitions and discussions of the examples may assist you to determine the level of development of your product when your agency has completed, or will have completed, its effort on the product's development.

Ready for critical review and for preparation for field test (prototype): a draft of the product has been completed and is ready for conceptual testing--the critical examination and review of the draft--and for putting the product into a format suitable for field testing; these latter activities, review and revision, will be undertaken by another agency.

Examples 1 and 5 describe products in this stage of development upon their release from the original agency. In example 1 the product has only been prepared in draft form; it is being released to another agency for critical review. Similarly, in example 5 the product is being forwarded to another agency for preparation for field testing. In general, then, products at this level of development have been drafted but are not yet in a form that can be used for large-scale field tests.

Ready for field test: the product has been modified as necessary based on the critical review and you have prepared a version that is clear and usable by members of the target population in a field test; the product may even have been through some field tests, but additional field testing will be conducted by another agency.

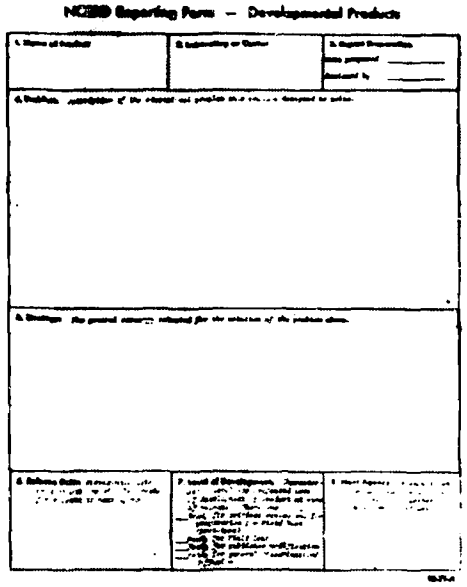
In example 2 the product is considered ready for field test because when you released the product it was in a form in which it could be tested. Although some tests had already been completed when you released the product, the product was not "ready for publisher modification" as additional developmental testing was planned.

Ready for publisher modification: the product has been field tested and you have revised it based on the results of the field test(s); the product is now ready for review by the publisher or dissemination agent prior to preparing it for publication; changes in content or format may be made by the publisher, as he is responsible for putting the product into its final form for dissemination. Many practical decisions are still to be made at this point. The publisher or distributor will make them, however.

Examples 3 and 6 describe products in this stage of development upon their release. In both of these instances, the publisher made additional, fairly major, modifications (beyond simple formatting and type-setting) in the product after you released it to him, so the product could not be considered ready for general dissemination when it left your hands.

Ready for general dissemination/diffusion: the product has been reviewed by the publisher or dissemination agent, if any, and you have made all the modifications necessary for preparing the product for dissemination; all that remains is for the product to be published and disseminated; this may already have occurred and the product may now be in use.

In example 4 the product is considered ready for general dissemination when you release it because all modifications of the project have been completed when you released it. Decisions about the method of reproduction may still be made, but the publisher will not make additional changes in the product itself.



SECTION 8. Next Agency

Instructions

Unless a product is to be released directly to users from your agency, you are likely to turn it over to a publisher, printer, or to another Laboratory or Center for additional developmental or marketing efforts. Give the name and address of the agency which received (or will receive) the product and become responsible for its development, marketing, etc.

It is important for you to give complete information about the agency to whom the product is being released for two reasons: first, to avoid confusion between agencies with similar names; second, to tell product evaluators where they can obtain the product after your agency has released it.

If your agency will release the product directly to users, or if you have not yet contracted with another agency for further development or for marketing, you should note this in this space.

Clear Examples

1.1 Copy-Print Centers
2206 S.W. Jefferson Street
Portland, Oregon 97201

2.1 Research for Better Schools, Inc.
1700 Market Street, Suite 1700
Philadelphia, Pennsylvania 19103

3.1 Technicon Educational Systems
590 East Middlefield Road
Mountain View, California 94040

Less Clear Examples

1.2 Copy-Print Center
Portland, Oregon

2.2 RBS

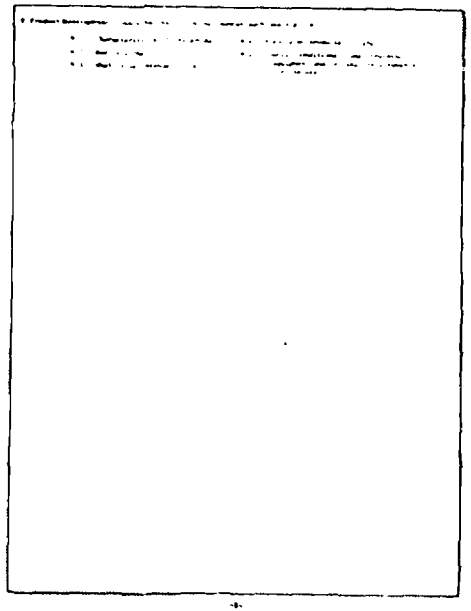
3.2 Jim Jones-Technicon

4.1 MacMillan Education Services, Inc. 8701 Wilshire Blvd. Beverly Hills, California 90211	4.2 MacMillan

Discussion of Examples

It is generally best to give the complete name and address of the agency. Often, as in examples 1.2 and 3.2, the agency will not be nationally known and the name and city may not be sufficient information to insure that an evaluator's inquiry reaches the agency. Similarly, as in example 2.2, an agency might be relatively well known, but the address may not be readily available to evaluators.

In example 4.2, it is possible that an evaluator would assume Macmillan Publishing Company was the designated agency, since it is better known than Macmillan Educational Services, Inc. Writing out the complete name and address of the agency will avoid such unnecessary confusion.



SECTION 9. Product Description

Give an overview of the product by describing the following:

1. Describe the characteristics of the product. Point out special features you think a reader should be aware of: for example, a product may be self-contained.
2. Tell how the product works, what procedures are followed in using the product.
3. Summarize the intended outcomes of the product, what you expect to happen if product is used.
4. Describe other products, if any, which are closely associated with this product, such as a teacher training program associated with a student's curriculum product.
5. Describe the special conditions, time, training, equipment, etc., required to use the product. The following are suggestions of factors to be considered:
   - time required to use the product
   - special staff training required
   - consultants, specialists (initial, ongoing)
   - special rooms, building, remodeling, etc.
   - equipment, machines, computer time sharing
   - reorganization of staffing pattern, new roles
   - paraprofessional aides, volunteers

Experience with product reporting has shown that there tend to be numerous omissions of important product information in this section, especially when existing documents are "cut and pasted" in an effort to provide the required information. This section is extremely important as a product overview for the reader, so great care should be taken to see that all points are covered.

If you do use descriptions from existing documents to fill out the form, rework them if necessary so that the information is in the format and sequence requested. It would be helpful for you to number the sections of your response that correspond to the five kinds of information requested. When the items are numbered they may be readily referred to by evaluators. The less clear example below shows how much harder it is to get a good overview of the product if these directions are not followed.

Some product characteristics that you may wish to include in your response to Part I are listed below.

- Skill oriented
- Multi-media
- Self-contained
- Nongraded
- Hierarchically sequenced
- Programmed sequence (by single frame or period)
- Computer-assisted
- Parent-community involvement
- Low reading level
- Performance objectives
- Behavioral objectives
- Student self-direction
- Process-oriented
- Inquiry or discovery method
- Interdisciplinary
- Bilingual
- Diagnostic and/or evaluative tests
- Individualization by:
  - Pace
  - Content
  - Degree of structure
  - Sensory modality/learning style
  - Process or method
  - Selection of activities

1.1

Clear Example

1. Characteristics of the Product. Minicourse Fifteen is a multimedia, self-contained, teacher education course, individualized by pace.

2. How it works. Each of five instructional sequences consists of three model lessons presented on videotape and readings in the teacher handbook. The handbook contains a chapter for each instructional sequence, as follows.

   Ch. 1: Independent Learning
   Ch. 2: Learning Contracts and Objectives
   Ch. 3: Independent Learning Contracts: Resources, Tasks, Deadlines
   Ch. 4: Independent Learning Contracts: Providing Reinforcers
   Ch. 5: Introducing Independent Learning in the Classroom

   For each instructional sequence, teachers read the appropriate handbook chapter, view the videotaped instructional lesson, plan a similar lesson, teach it to a small group while being videotaped, and evaluate their performance. Most of the instructional sequences also have provision for reteaching and re-evaluation of the lessons. (This process of viewing models and then practicing the skills yourself while being videotaped is known as microteaching.)

3. What it is intended to do. Participation in the minicourse prepares teachers to introduce the students to skills for independent learning. As part of their microteaching experience the teachers systematically introduce their students to independent study and start two or more students on independent projects. By the end of the course, teachers should be able to include independent study as an integral part of their regular instructional program.

4. Associated Products. There are two related minicourses. Minicourse Eight, "Organizing the Primary Classroom for Independent Learning and Small Group Instruction," is awaiting commercial release by Macmillan Educational Services, Inc. Minicourse Eleven, "Teaching Skills that Develop Independent Learning in the Secondary Classroom," is scheduled for preliminary field testing in November, 1971.

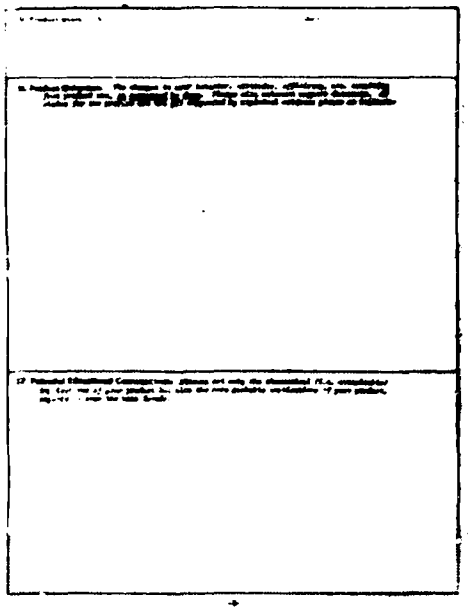
5. Requirements for Use:
   1. A videotape recorder, monitor, camera, and microphone. This equipment will service up to 10 teachers at a time.
   2. A room for viewing and conducting the videotaped microteach lessons.
   3. Two released class hours per week per teacher for microteaching lessons, plus one after-school hour for viewing the instructional tape. In addition, the teacher will spend 1-1 1/2 hours in preparation for each of the two weekly lessons.

1.2

Less Clear Example

The course consists of five lessons in which the teacher practices approximately three or four skills. Each lesson includes a videotaped presentation of a model performance and a chapter in a handbook that presents a general discussion of the skills. The handbook provides general directions for the conduct of the courses, written exercises that supplement the videotaped presentation, and evaluation checklists for the teacher to use in assessing her recorded performance. The teacher practices the skills in each lesson with a small number of students each week. Her performance is recorded. After playback and evaluation of her initial try, she then practices the same set of skills a second time. The second practice session may occur with a small group or with her regular class. Each lesson has a different requirement in this regard.

SECTION 10. Product Users



Instructions

Products are usually designed to be used by a specific group, such as elementary school science teachers, urban school principals, primary grade students, and so forth.

In this section you should describe the individuals or groups who are expected to use this product. First, select appropriate descriptors from those listed at the bottom of this page. Select as few or as many as necessary to characterize the users. You should be as explicit as possible in describing the intended users of the product. Second, estimate the total number of persons who conceivably could use the product, that is, the total possible "market." This number should be an estimate of the population who could eventually use the product, not just the number who have already used the product during its development, such as in the field testing.

It is possible that your product may be used by one group (such as teachers) for another group (such as children). If this is the case you should be sure to describe the users of the product, not the beneficiaries. You may, if you wish, provide some information about the beneficiary in addition to the description of the user.

Select as many of the descriptions below as you wish to describe your intended product users. You may use any other descriptions you wish, as well. Note: it is not necessary for you to select a description from each category, though you should include descriptions from any categories that are appropriate.

1. Demographic Characteristics

   Rural, Suburban, Urban, Reservation, Inner City

2. Family Income Level

   Lower, Middle, Above Middle

3. Educational Level

   Early Childhood (Nursery, Preschool)
   Elementary (Grades 1-6)
   Secondary (Grades 7-12)
   Higher Education (Undergraduate and Graduate)
   Adult Education

4. Ethnic Characteristics

   American Indians
   Latin-Americans
   Afro-Americans
   Caucasians
   Asian-Americans

5. Educational Personnel

   Policymakers (School Board Members)
   Administrative Personnel (Superintendents, Principals, and Supervisors)
   Instructional Personnel
      Teachers (Subject area: Math, Science, etc.)
      Special Education Teachers
      Teacher Trainees (Preservice, Inservice)
      Teacher Trainers
      Tutors
   School Support Personnel
      Instructional Media Personnel
      Information System Specialists
      Physical/Speech/Hearing Therapists
      Guidance Counselors/Psychologists
      Educational Consultants
      Teacher Aides
      Parent Community Volunteers
   Research & Development, or Related, Personnel (Evaluation Specialists, Curriculum Specialists, Etc.)

6. Geographic Area

   State-wide application (Alaska, Arizona, etc.)
   Regional application (Northwest, Southwest, etc.)
   National application

7. Physical and Mental Characteristics

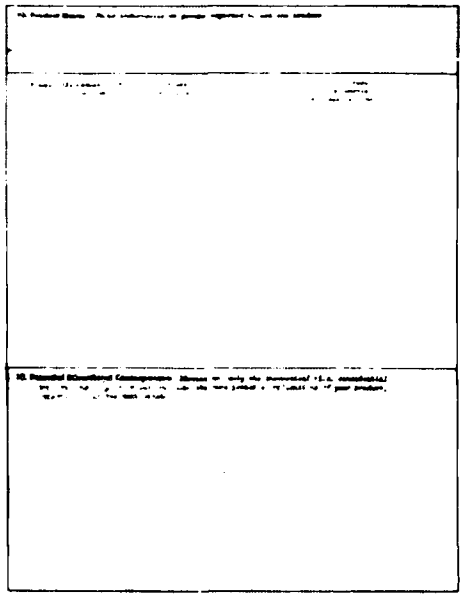
   Slow Learners/Retarded Children
   Physically Handicapped (Blind, Deaf, Etc.)
   Emotionally Handicapped
   "Learning Disabilities"
   "Educationally Disadvantaged"

8. Linguistic Characteristics

   Bilingual
   Non-English Speaking
   Dialect Speaking

Clear Examples	Less Clear Examples
1.1 The product is to be used by teachers of bilingual children in grades 4-6 who come from low-income families in the rural Southwest. Estimated total number of rural Southwest teachers of grades 4-6: 700 (for 20 000 children).	1.2 30 teachers used the product in the field test; 50 schools will be using the product next year.
2.1 The product may be used by elementary school principals, in all geographic regions, interested in improving teacher classroom effectiveness. OE statistics indicate there were 67,000 elementary school principals in the US in 1968; of these perhaps 5%, or 3350, might be interested.	2.2 This product can be used profitably by all administrative personnel in the country.
3.1 Secondary school teachers, in northern urban settings, who teach children from lower income families in recently desegregated areas. There is no authoritative estimate of the number of such teachers. Agency staff, however, estimates there must be at least 5,000-10,000, perhaps even as many as 30,000, such teachers.	3.2 Secondary school teachers in urban settings.
4.1 The product is for Indian preschool children, on reservations or in rural areas, who lack proficiency in English language skills. The Bureau of Indian Affairs, in 1970, estimated that there were 18,000 such children.	4.2 Indian children.

SECTION 11. Product Outcomes



Instructions

In this section you should discuss the results or effects of the product. Specifically, you should describe what the users do differently as a result of your product. What changes in knowledge (content areas) or skills (cognitive, psychomotor, social) or what affective changes can be attributed to your product?

In addition to specifying the area of change, you should also report the amount of change that occurred and the measures used to assess these changes. You should clearly identify all support documents you have describing these outcomes in more detail. Support documents may be reports of field tests or other evaluation studies, data from tests of the product after it has been disseminated, or other documents containing information about the product's outcomes.

If a product is in the prototype stage and has not yet been tested, this fact should be noted. In this case you may wish to describe the outcomes you expect will occur.

Clear Examples

1.1 A feasibility study indicated that children's self-concept, as measured by Osgood's semantic differential technique, changed significantly in a positive direction ($p<.01$) after using a similar product. Our product is only in draft form now, so no data is available about its outcomes, but we expect our product to have similar effects.

2.1 The final product has not yet been tested but 30 children using the prototype product were observed (using BZT techniques) to decrease incidences of disruptive behavior an average of 50% and to increase attending behaviors an average of 30%. The present product is expected to have greater gains since it has been revised.

3.1 Eighth grade children achieved an average gain of 1.4 years (grade equivalent) in word meaning scores on the SAT, in an

Less Clear Examples

1.2 Self-concept gains were found with similar products.

2.2 Desired changes in pupil behavior were achieved using the prototype form.

3.2 Eighth grade children's scores in word meaning improved significantly over an 8 month period.

8 month period. (See R&D Center Technical Report #8-201.)

4.1 Product field test reports showed that teacher's use of inquiry techniques in science classes increased 80% over their base rate after using the product for five sessions. (See R&D Center Technical Report #8-201.)

Studies comparing attitudes toward school work of students in classes where teachers had used the product were significantly more favorable than attitudes of students from other classes ($p<.05$). (See "Process Oriented Education -- its Effects on Student Attitudes," ERIC document #ED 002 714.)

Use of the product in an undergraduate teacher training program resulted in 1) a 50% increase in use of inquiry skills in practice teaching and 2) more favorable ratings by trainees' field supervisors of trainees' classroom effectiveness, as compared to attitudes toward other trainees not using the product. Results based on data from interviews with trainees and their field supervisors (interviewers used the critical incident and card sort techniques). (Nelson, A.M. "New Approach to Teacher Training," Journal of Ed.Psych, vol. 13 No. 4, 1970.)

4.2 Teachers improved in the use of inquiry techniques after taking these materials; students liked the classes better, and teacher trainees using the materials both improved in the use of inquiry skills and were more effective in the in the classroom.

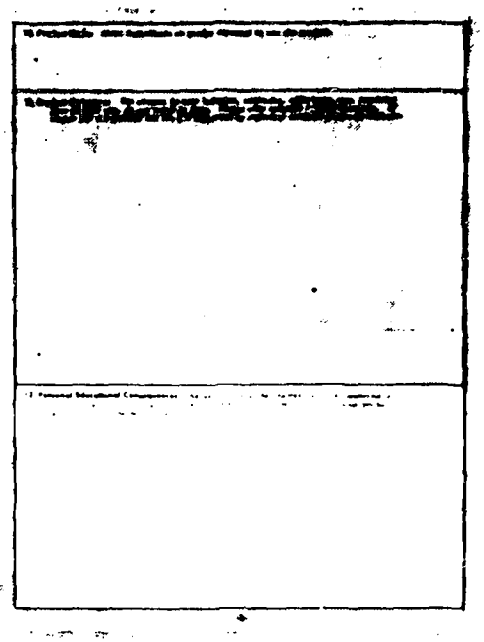
Discussion of Examples

Because the outcomes of a product help determine the usefulness of the product, it is important that they be stated clearly and concisely. You should be able to cite evidence regarding the outcomes that have occurred (examples 3.1 and 4.1), or evidence supporting your expectations that certain outcomes will occur (examples 1.1 and 2.1).

Lack of support or insufficient explsnation of the product outcomes, as seen in the Less Clear Examples, may give rise to questions such as the following:

1. How do you know it wouldn't have happened anyway?
2. How did you determine what changes occurred?
3. How do you know your product is responsible for the change?

Asking yourself these questions about your product outcomes will help you determine whether or not your statement is complete.



SECTION 12. Potential Educational Consequences

Instructions

Even though a product has had, as yet, only limited impact, it may have far-reaching potential educational consequences. For example, a successful computer managed system of instruction may not have much impact at this time because most school systems cannot afford, or do not have access to, the requisite computer facilities. However, if such a system were implemented on a large scale, it is conceivable that even the physical structure of schools would change. There would be less need for large classrooms and more need for small seminar rooms, terminals, and individual study rooms. A more probable consequence, even with only limited adoption, of a computer-managed instructional system would be an increase in the degree to which education is individualized. Each student would study units whose content and format are most appropriate for his interests, abilities, and goals.

In this section you should state the potential educational consequences of your product, both the possible and the probable. The theoretical, or possible, consequences are those effects which could conceivably occur as a result of your product's use. Probable consequences are those effects which you are reasonably certain will occur if your product is used. Note that there is a distinction between these consequences and the outcomes discussed in the previous section. "Outcomes" are the immediate effects of a product's use and are generally defined as changes in the behavior of the users. "Consequences" are longer-term, generalized effects which extend beyond the group of product users. Thus, this section provides an opportunity for product developers to convey the potential significance of the product apart from the results gained to date. You should be sure that the relationship between your product and the potential consequences is clear. In other words, you should explain why you expect that your product will have certain consequences.

Clear Examples

1.1 It is anticipated this reading program could greatly enhance the reading skills of minority group children. The cost is very moderate ($3.50 per child per year) and disadvantage. students in primary grades who used the program showed reading gains 50% greater than students in a control group. In addition, it is likely that the general academic performance of students using this product will improve because doing well in other subjects usually requires good reading skills. This general academic improvement may, in turn, improve students' attitudes toward school and toward education in general.

Because the program utilizes commonly available materials, such as magazines, rather than textbooks, there will likely be a substantial financial savings to school district. And, because the reading program incorporates peer tutoring, it should take less of the teacher's time, enabling her to spend more time working with students in areas where her skills are required.

2.1 Students who have participated in the Career Education game when they entered high school will most likely make more satisfying

Less Clear Examples

1.2 Reading problems of disadvantaged children will be solved; students will like school more; expenses for materials will decrease; and, teachers will have more time to spend working with students.

2.2 The Career Education game will have a substantial impact on decisions students make, and it may have implications for the role of the counselor as well.

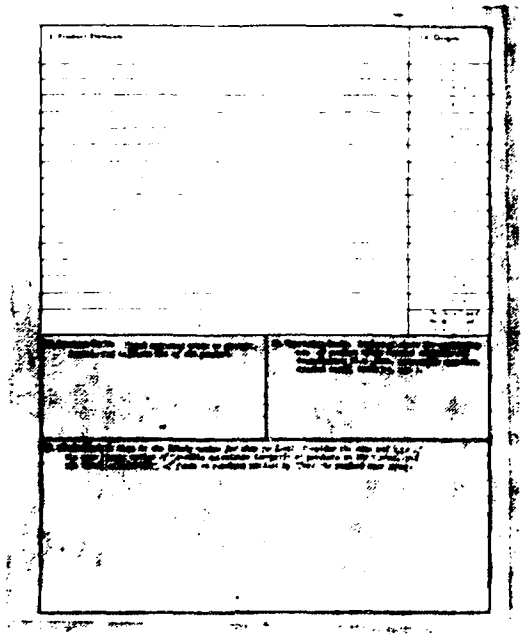
decisions about which courses they will take in high school and about what they will do after high school. While the objectives of the game are to introduce students to the decision-making process and to procedures for acquiring, considering, and evaluating information in making decisions, the students are making the same decisions for hypothetical students that they must later make for themselves; the game, then, not only provides them with practice in decision making, but also allows them to examine the consequences of their decisions in a "safe" context. It is also probable that these students will take more steps to identify and explore various post-high school options because they will have been made aware of the need for actively deciding what to do before graduation is upon them.

An additional possible consequence of this product would be the generalization of the effects to students who did not participate in the game. If enough interest is generated in those students who were involved in the game, it is likely to spread to their peers.

Discussion of Examples

In this section you have an opportunity to discuss the potential contributions of your product beyond the expected outcomes discussed in the previous section. Examples 1.1 and 2.1 clearly specify both possible and probable potential consequences of the product. In addition, they briefly explain why they believe these consequences might follow use of the product.

Example 1.2 mentions briefly some potential consequences of the product, but these statements are not developed at all, so a reader cannot tell how or why these consequences may occur. Example 2.2 does not even specify consequences that may result from the product; it simply asserts that there will be an impact. Both of these examples provide very little information, especially when compared with examples 1.1 and 2.1.



SECTION 13. Product Elements
SECTION 14. Origin

Instructions

SECTION 13. PRODUCT ELEMENTS

List the elements which constitute the product. The elements are physical entities (such as handbooks, filmstrips, tests, workbooks, tapes, scoring keys, computer programs, and so forth) which make up the total product. Research studies, field test reports, position papers, and the like are considered support documents, rather than product elements, and should not be listed in this section. Support documents should be cited to support or elaborate on points made in response to other questions such as problem statement, expected outcomes statements, etc.

This list should tell, by name, everything which constitutes the total product. Give the titles of all books, videotapes, films, etc. that are part of the product. If the list is very extensive, you may continue on a separate attached list giving the titles of the product elements.

SECTION 14. ORIGIN

For each element indicate its origin by circling the appropriate letter:

D, if it was developed by your agency;
M, if it was material modified from another source but modified by your agency for this project;
A, if it was adopted from another source.

See the definitions below for elaboration of these terms.

Clear Examples

SECTION 13. PRODUCT ELEMENTS AND ORIGIN

1.1 Teacher's Guide: Coordinated Helps in Language Development (CHILD) (D) M A

2.1 11 Instructional and Model Films (See attached sheet for titles). D M (A)

Teacher's Handbook: Effective Questioning, Elementary Level (D) M A

Coordinator's Handbook: Effective Questioning, Elementary Level (D) M A

A Microteaching Approach to Teacher Education (a book of professional research) D (M) A

The Minicourse in Your School (an information handbook) (D) M A

Less Clear Examples

1.2 Teacher's Guide: CHILD (D) M A

2.2 11 Films (D) M (A)

Teacher Handbook D M A

Coordinator's Handbook D M A

Professional Research Book D M A

Information Handbook D M A

3.1 Curriculum Guidelines for Inner-City Teachers	D M A
SRA Teaching Problem Lab:	
5 Filmstrips/Tapes (see attached list)	D M (A)
3 Films (Kinescopes of teaching situations) (see attached list)	D M (A)
2 Demonstration videotapes (see attached list)	(D) M A
Assorted Testing Materials (see attached list)	D (M) A
Classroom Climate (Teacher Workbook)	(D) M A

3.2 The Innovation Team Summary Report	(D) M A
Evaluation Report CITL 1969	D (M) A
Follow-up study of 1967-68 CITL Graduates	D (M) A
Summary of CITL 1968-69 MCREL	D (M) A

Discussion of Examples

SECTION 13. Product Elements

Example 1.1 shows that titles, rather than generic categories of instructional materials should be given as product elements. Examples 2.1 and 3.1 show that for more complex products, it may be necessary to attach lists of titles to supplement the Product Elements Section. The films, filmstrips, books, etc. adopted as part of the product may be very numerous and their titles can therefore be appended.

Examples 1.2 and 2.2, on the otherhand, demonstrate incomplete identifications of the product elements. Example 3.2 lists evaluation and summary reports, rather than product elements. These should be listed as support documents rather than elements used when the product is in operation.

In this section, then, you should list only those elements which the individuals or groups using the product will need.

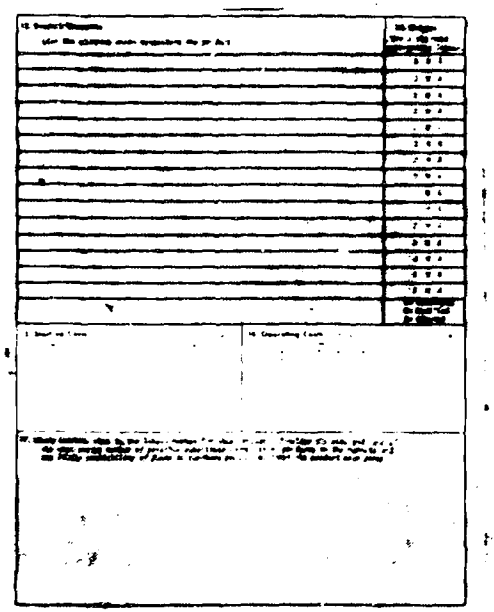
SECTION 14. Origin

D = Developed, i.e., elements which your agency was responsible for developing.

M = Modified, i.e., elements which another individual or agency developed but which your agency modified, or revised somewhat, for use with this product.

A = Adopted, i.e., elements which were developed by another individual or agency and which are being used with this product without being modified or revised in any way.

SECTION 15. Start-Up Costs
SECTION 16. Operating Costs



Instructions

SECTION 15. START-UP COSTS

In this section state the total costs to procure, install and initiate use of the product. Include costs for purchase of special equipment required to use the product. Indicate the "hidden" costs, if any, associated with the product, such as required building remodeling, released time for teacher training, etc.

SECTION 16. OPERATING COSTS

In this section give the costs to use the product after installation. Include costs for fees, consumable supplies, special staff salary, etc. If possible, estimate these costs in units per pupil/teacher/etc.

Clear Examples

1.1 Course Purchase: $1,400
Course Rental: $200 per 8 teachers
Videotape equipment: $3,000 (approx.)

2.1 Leader's Guide $6
Participant Materials $2.25 (each)
Audiotape $3.25
Time to train Instructor: 23 hours
(No additional costs to operate)

Less Clear Examples

SECTION 15. START-UP COSTS

1.2 $1,400

2.2 $11.50

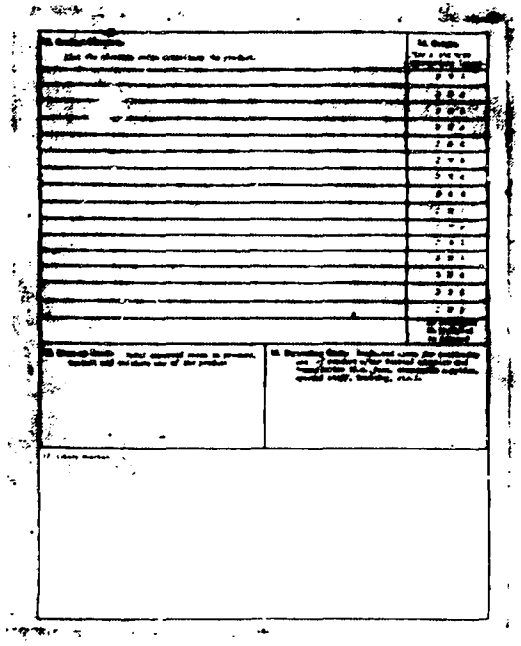
SECTION 16. OPERATING COSTS

1.1 For each teacher: 11 hours of substitute teacher time (approx. $60)	1.2 Substitute Teacher Time
2.1 20 hours of Instructor time for workshop for 30 to 36 participants 35 hours of class and study time for each teacher attending workshop	2.2 Instructor Time

Discussion of Examples

As you can see, the clear examples give not only a figure of estimated total costs, but some general breakdown of what must be purchased to use the product, services required, and cost per participant. The less clear examples do not give breakdowns of the total costs, and, in the case of operating costs, they do not even give estimates in money terms.

Each product should have an estimate of the cost to adopt the product (Start-Up Costs); for some products, however, there may be no ongoing operating costs. All products, then, should have some estimate of cost, even if the product is not yet completed. If firm figures are not available, you should give the best estimate you can make.



SECTION 17. Likely Market

Instructions

In this section you should describe the likely market for your product. In your description you should include information on the size and characteristics of the user group, the number of products already available that will be in competition with yours, and the amount of money users are likely to have available to purchase the product.

If you have data from formal market analyses available you should present it here and indicate the source. Otherwise you should give your best estimates of the information requested. Finally, if your product has been available to users, you should report how widely it has been adopted, and by whom.

Clear Examples

1.1 A needs assessment was carried out by an independent research firm to determine the number of major school district administrators who claim they would adopt the product if it were available now. One hundred and eighty administrators from sixty districts in three states were surveyed; 80% indicated that they would favor adoption of the product at its present cost. (See Market Analysis Report, Support Document # 17, attached.) Two other, similar, products are also available, and one is being marketed at a lower price, so the 80% estimate may be somewhat high for purposes of generalizing. If 60% of the districts in the three states adopted the product, it would come to nearly 1200 districts.

2.1 Anticipation of greater production efficiency (learning/dollar invested in instruction) has led two-thirds of the state legislatures in the United States to require or strongly encourage adoption of a comprehensive management and planning system by schools (see Support Document # 2). Since this package is the only product available for training consultants and school personnel to adopt systematic management and planning procedures, the potential market

Less Clear Examples

1.2 A needs assessment yielded favorable results. See Support Document # 17.

2.2 Review of the literature, requests from the field, and legal requirements indicate a strong demand for a comprehensive management system.

for the product is very broad. And because such a system will be required of schools, funds must be made available for its purchase.

3.1 Approximately 3500 students enroll in teacher training programs each year in California alone. In addition, 65% of the state's school districts engage in inservice training activities for their teachers. Thus the number of potential consumers is large. Also, this is the only packaged product which endeavors to train teachers through a practice and feedback oriented approach. However, most school districts and training programs do not have funds available to purchase our product outright. As a result, the product is being marketed chiefly through rentals rather than sales. In the past year, 57 institutions rented the package and requests were received from an additional 35 groups.

3.2 There are a large number of trainees who could benefit from our product, and there are no similar products available. Many institutions, however, have not been able to afford to buy our product, but they have rented it.

Discussion of Examples

The need or demand for a product may be established in several different ways; whatever the technique employed, it should be briefly described. Note that references can be given to support documents which elaborate on the results and procedures of the market analysis. However, you should summarize the findings on the form, as in examples 1.1 and 2.1, rather than simply noting the reference to the support document, as in 1.2 and 2.2. Example 3.2 provides the kinds of information requested, but it lacks detail and, as such, is rather ambiguous. Example 3.1 provides either statistics or estimates to delineate more precisely the extent of the market. Note how this example also provides information on the actual use of the product, as well as the likely market.

THANK YOU FOR YOUR EFFORT IN PRODUCT DEVELOPMENT

Product Reporting Form — Developmental Products

1. Name of Product	2. Laboratory or Center	3. Report Preparation *Date prepared* _____ *Reviewed by* _____

4. Problem: *Description of the educational problem this product designed to solve.*

5. Strategy: *The general strategy selected for the solution of the problem above.*

6. Release Date: *Approximate date product was (or will be) ready for release to next agency.*	7. Level of Development: *Characteristic level (or projected level) of development of product at time of release. Check one.* ___ *Ready for critical review and for preparation for Field Test (i.e. prototype materials)* ___ *Ready for Field Test* ___ *Ready for publisher modification* ___ *Ready for general dissemination/diffusion*	8. Next Agency: *Agency to whom product was (or will be) released for further development/diffusion.*

10-71-A (D)

9. Product Description: *Describe the following; number each description.*

- *1. Characteristics of the product.*
- *2. How it works.*
- *3. What it is intended to do.*
- *4. Associated products, if any.*
- *5. Special conditions, time, [illegible], equipment and/or other requirements for its use.*

10. Product Users: *Those individuals or groups expected to use the product.*

11. Product Outcomes: *The changes in user behavior, attitudes, efficiency, etc. resulting from product use, as supported by data. Please cite relevant support documents. If claims for the product are not yet supported by empirical evidence, please so indicate.*

12. Potential Educational Consequences: *Discuss not only the theoretical (i.e. conceivable) implications of your product but also the more practical implications of your product, especially over the next decade.*

13. **Product Elements:** *List the elements which constitute the product.*	14. **Origin:** *Circle the most appropriate letter.*
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D= Developed M= Modified A= Adopted

15. Start-up Costs: *Total expected costs to procure, install and initiate use of the product.*

16. Operating Costs: *Projected costs for continuing use of product after initial adoption and installation (i.e.,fees, consumable supplies, special staff, training, etc.).*

17. Likely Market: *What is the likely market for this product? Consider the size and type of the user group; number of possible substitute (competitor) products on the market; and the likely availability of funds to purchase product by (for) the product user group.*

Instructions For Completion Of

Product Reporting Forms -

Knowledge Products

FORM 10-71-B

General Information

- WHAT IS A KNOWLEDGE PRODUCT?

A product is a solution to an educational problem. By definition, then, a knowledge product fills an important gap in our knowledge about subjects or topics relevant to education. The generation of that new information should permit major progress to be made in either basic or applied activities; progress which would not have been possible without the creation of that new product. For example, a knowledge product may provide new information about effective learning strategies for elementary school children; or it may contribute new knowledge concerning more effective school management techniques; or it may provide data concerning the effectiveness of certain instructional programs. Regardless of the area of focus, however, the new knowledge product does not become a "product" until it is readily available to other educational practitioners. Typically, this availability is made possible through a research report, journal article, monograph, or some other form of semi-permanent, retrievable, mass communication.

- WHAT ARE THE VARIOUS TYPES OF KNOWLEDGE PRODUCTS?

There are five basic types of knowledge products:

(1) *Literature reviews*: reviews of existing knowledge summarized along lines not previously available;

(2) *Reports of research*: reports of studies designed to test educational hypotheses, investigate problems, or discover basic relationships;

(3) *Theoretical papers/research syntheses*: analyses of existing research leading to the development of new insights, theories, or conceptualizations;

(4) *Papers on model designs/specifications*: designs and descriptions of the component parts, and interfaces among the parts, of an educational system, or a model for producing educational change; and

(5) *Evaluation or feasibility studies:* analyses of educational projects, or proposed projects, to assess their effectiveness, or feasibility, in terms of specified criteria.

- WHAT IF MY PRODUCT MAY BE CLASSIFIED AS TWO OR MORE OF THESE TYPES?

In reality, a product may involve a combination of characteristics. A systems analysis of urban education might also include a literature review and an evaluation of existing urban education projects. Similarly, an evaluation of a specific individualized instruction program might also include an analytical synthesis of the findings of other, similar evaluation projects. Each should be classified in terms of its primary emphasis, however.

- HOW MANY FORMS DO I FILL OUT?

Complete one form for each significant new product that you have developed. If you have been engaged in the programmatic investigation of a problem area, you may have produced a series of conceptually or methodologically related products. If each presented new findings, they should all be reported separately, but they should be cross-indexed to each other (in Sections 3, 4, 5). In deciding which products to report, keep in mind that a new product must provide new knowledge relevant to education.

- WHY NOT JUST SEND OUR PUBLICATIONS LIST?

This aspect of the NCERD Products/Accomplishments Reporting System is concerned only with new knowledge products. Some publications are an effort to communicate the same basic information to a variety of different audiences; other attempt to expand total exposure. This latter is especially the case when several journal articles are produced which report on subsections of a larger study reported, and available in, say, OE Final Report Form. Also publication lists frequently include brocures, newsletters, posters, and other public relations documents. Thus a publication list usually goes far beyond listing only "new knowledge" reports.

- CAN I SEND PARAGRAPHS AND PAGES FROM OTHER OF OUR DOCUMENTS?

It would generally be quite unwise to do so. As you complete this form, you will be asked to observe fairly specific instructions related to each question. Abbreviated examples will illustrate these instructions. It is highly unlikely that "cutting and pasting" from pamphlets, brochures, annual reports, etc., would respond specifically to the instructions.

If, however, you feel that a particular document provides additional support or elaboration for the material you have written, you are encouraged to cite it as a support document. Support documents, i.e., documents providing additional support for, or explanation of, your product, are especially desirable in the general problem, strategy, and implication areas. When you cite support documents be sure to identify them completely.

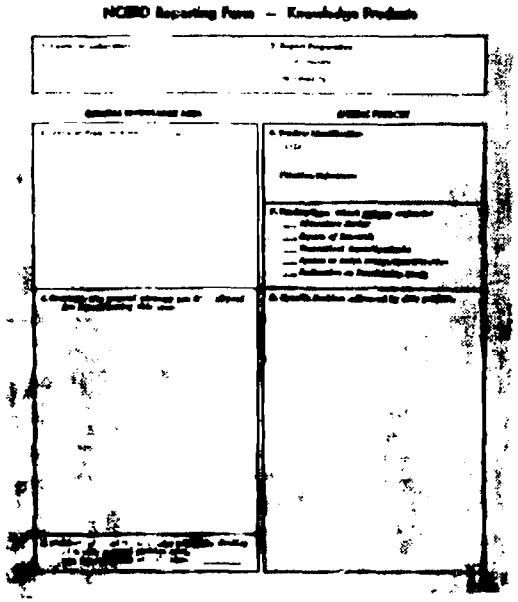
- WHAT IF I NEED MORE SPACE FOR MY COMMENTS?

If you need more space than is provided on the form, continue your answer on a separate page and then attach it to the form. Be sure to indicate the extension of your response by writing "continued on attached page" at the end of that portion of your response recorded on the form.

- - - - - - -

If you have any questions concerning the completion of this form, please call the Product Coordinator at (415) 328-3550, ext. 900.

SECTION 1. Center or Laboratory
SECTION 2. Report Preparation
SECTION 3. General Problem Area



Instructions

SECTION 1. CENTER OR LABORATORY

Enter the name of the laboratory or center responsible for the development of this product.

SECTION 2. REPORT PREPARATION

Date prepared: Enter the date that this product reporting form was completed.

Reviewed by: Upon review of the completed product report, the laboratory or center director (or his designee) should sign the form indicating it has been reviewed for accuracy and released for transmission to OE.

SECTION 3. GENERAL PROBLEM AREA

In this section you should describe the general problem area, or conceptual context, with which your product is concerned. Because this form is concerned with knowledge products, the problem being dealt with will most likely be a lack of knowledge or information about a general topic or topic area.

In describing the general problem area it is important to define it at the appropriate level of specificity. In general, it is expected that a general problem area will be comprised of a number of smaller, more specific, problems; that is, a number of separate knowledge products will be required for the solution of a general problem. For example, a number of separate research studies would be required to resolve the problem of *"a need for information on characteristics of the learning environment that contributes to primary-grade children's ability to manage their own behavior in class."* If a problem is defined such that a single knowledge product "solves" the problem, then a specific problem, rather than a general problem area, has been stated. Statements of specific problems are requested in Section 8 and should not be given here.

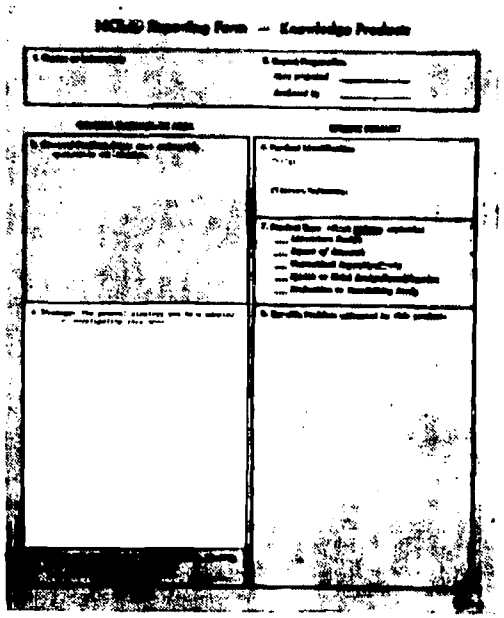
The specific problems comprising a general problem area should be sufficiently homogeneous in topic to warrant their inclusion as aspects of the superordinate *"general"* problem. If you defined a general problem area as *"a need for information regarding effective techniques in in-service training for teachers of bilingual students,"* then all the specific problems comprising this general area would relate to the study and evaluation of techniques for training teachers of bilingual students. You should avoid defining the problem area too broadly. *"A need for information about procedures for individualizing education"* would include problems relating to information about physical facilities and resources required for an individualized educational program, about alternative models of the role of the teacher in individualized education, about methods of organizing and structuring the curriculum to accommodate individualization, and so forth; each of these would themselves constitute general problem areas, and thus, should not be viewed as specific problems of individualizing education. In general, then, a general problem area should be defined at a level at which it may be comprised of a series of homogeneous, specific problems.

The purpose of defining the general problem area in which your product falls is to assist a reader in identifying the general context of your work. It is not necessary that your agency be responsible for developing all of the specific products necessary to resolve the general problem. It is necessary only that your products contribute to the solution of the general problem. You may, in fact, have only one specific product relating to the general problem area.

After you have fully described the general problem area for the first of the specific products relating to that area, you need only briefly identify it for subsequent products in the same area.

Clear Examples	Less Clear Examples
SECTION 3. GENERAL PROBLEM AREA	
1.1 We are attempting to identify specific factors responsible for low reading performance in bilingual students and develop procedures which will improve their reading.	1.2 Teachers need information about the availability of social studies materials for use with secondary Chicano students. (This problem is too narrowly defined.)
2.1 Information is needed to define the relationship between visual perception and the development of reading skills in children in primary grades.	2.2 Information is needed about the relationship between perception and learning. (This problem is too broadly defined.)
3.1 Objective information about the nature of the teaching job and about the characteristics of people going into teaching is needed by teacher educators to assist them in identifying trainees likely to become committed to the teaching profession and to assist them in developing teacher training programs.	3.2 We need to know more about teachers and about their jobs. (This statement is too vague.)
4.1 Our agency is attempting to develop a comprehensive model for urban learning systems which will cover all relevant aspects of the system including communication patterns and channels of authority. This model should facilitate the analysis and improvement of urban schools.	4.2 A model for educational systems needs to be developed that can also be applied to urban learning systems by school planners. (This statement is also unclear.)
5.1 Information about the effects of the Athens Plan needs to be reported and made available to professional educational researchers and to school systems.	5.2 There is little public knowledge about the effects of the Athens Plan. We plan to fill this gap. (This statement should be more precise.)

SECTION 4. Strategy



Instructions

In this section you should outline the general strategy you have adopted, or are adopting, in searching for a solution to the general problem described in Section 3. You should describe as clearly as possible the approach you are taking, and why you selected that approach over other alternatives. You may also want to include a brief description of other strategies that were considered. You should include in this discussion both the operating guidelines you are following, and a description of the domain of responsibility being assumed by your agency for the problem's solution. If other agencies are to be involved with you in solving the problem, their responsibilities should also be described here.

A statement of your general strategy will permit a reader to more adequately understand how the specific products you have produced relate to the general problem described in Section 3.

This section is <u>not</u> intended as an overview of the procedures used in developing the individual product; information on the specific methods involved should be presented in Section 9. Rather, this section is intended to provide a description of the <u>general</u> approach taken in providing a solution to the general problem and reasons why this approach was adopted.

If you are reporting here on the first of a <u>group</u> of products relating to the <u>same</u> general problem area, your discussion of the strategy should cover the approach being taken for <u>all</u> the products relating to the general problem. In reporting subsequent products related to this general area you will not need to describe the strategy again.

Clear Examples

1.1 The professional literature was reviewed to 1) determine what theories have been developed and studies conducted about the relationship of bilingualism to reading performance and 2) identify those variables which seem to have the most harmful effect on reading performance. Studies were then conducted to explore the effect

Less Clear Examples

1.2 We have prepared a review of the literature relating to the relationship between bilingualism and reading. The review covers 653 references and, in addition to describing what has been done in this area, contains a summary of what factors seemed to have the most harmful influence on reading performance. (This statement

of each factor on reading and to test procedures for ameliorating the effects of each variable. The results of this effort were synthesized into a position paper containing a set of recommendations for teachers of bilingual students and for curriculum developers.

2.1 A series of experimental studies will be conducted, each of which analyzes the development of one or two specific aspects of visual perception and the role of these perceptual abilities in the development of reading skills. Studying specific aspects of perception rather than visual perception in general will provide more precise information about the role of perception in learning how to read.

3.1 Our agency is only engaged in the identification and validation of factors contributing to a teachers' commitment to the teaching profession. Our strategy is first to identify potential factors through an analysis and synthesis of existing literature, then to observe a sample of teachers from across the country in an effort to confirm the importance of the various factors. By beginning with a research synthesis we hope to eliminate the need for initial surveys. The observation of teacher performance will tend to keep our work empirically oriented.

4.1 Relevant dimensions of urban learning systems were identified through a factor analysis of responses of system personnel to questionnaires and interviews. An interactive process of model development and testing by comparison with observed events was adopted to develop a model for each dimension. This approach was selected over one involving computer simulation because of the greatly increased costs for little gain of the latter strategy.

5.1 A series of small studies was chosen to a) afford greater control over data collection, b) allow for revision of procedures based on experience, and c) minimize time demands on staff and students. The studies were conducted to explore the effects of the Athens Plan on academic performance, interpersonal relations, social behaviors, and post-high school experiences. Both alumni and current students were sampled to control for effects of time.

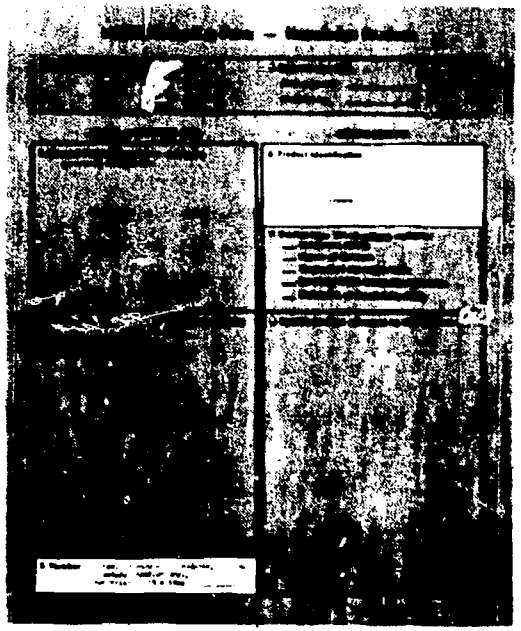
describes the product, not the strategy followed in seeking a solution to the general problem.)

2.2 Experimental and correlational studies will be conducted to analyze the development of specific aspects of visual perception and the role of these perceptual abilities in the development of reading skills. (This statement does not explain why this strategy was selected.)

3.2 Our agency will review the literature and then interview some teachers to discuss the results of our literature review. (This statement is too vague.)

4.2 We have been asked to assist the state department of education by developing a model of the state educational system. (This statement describes the objective, not the strategy.)

5.2 Studies include the effect of Athens Plan on academic performance, social behaviors of Athens Plans students, and effects of Athens Plan on student's post-high school plans and activities. (This statement identifies various products rather than describing the strategy.)



SECTION 5. Number

SECTION 6. Product Identification

Instructions

SECTION 5. NUMBER

If the specific product you are reporting here is only one part of your agency's effort in this problem area, and if you are reporting other products related to this general problem area, these other products are considered "associated products." For example, if you were reporting an evaluation study, associated products might be a comprehensive literature review, several specific research studies, a synthesis of the results of these studies, and perhaps even development of a new instructional model. In this case you would have developed a total of five specific products related to the general problem area. If you are reporting all five of these associated products at this time, you would indicate this by putting a 5 on the line in Box 5. If you are only reporting three of these associated products, you would write a 3 on the line.

SECTION 6. PRODUCT IDENTIFICATION

As was mentioned in the introductory definition of a knowledge product, it must be in a form which is readily available and retrievable. In the vast majority of cases, this form will be a written document. Papers presented orally at professional meetings and not retrievable via a service such as ERIC would not be considered finished products. You may, however, report a product as in process if you expect it to eventually be published. To help identify your product, list the title (or the anticipated title) of this document and other citation information in this box. Citation information might include:

(a) Authors;
(b) Publication date (or estimated release date);
(c) Other bibliographic data such as source (journal, volume, and page; publisher and copyright date, etc.)

If your product is in some A-V form other than a written document, you should provide any information which will help reviewers to identify it.

Note: In identifying your agency's products you should be sure to list only those which present new knowledge. For example, if the "product" is a computerized file of the Education Index, this should not be reported; it simply represents a new form of packaging of the Education Index. If, however, you have prepared an extensive review of the literature in an area that hasn't been reviewed before, then this should be reported because this does represent new knowledge.

Clear Examples

Less Clear Examples

SECTION 6. PRODUCT IDENTIFICATION

1.1 The Relationship Between Bilingualism and Reading: A Review of the Literature. T. Thompson. (Technical Report #87-158, Feb. 1971).

1.2 A review of the literature on the relationship between bilingualism and reading.

2.1 "An Analysis of the Degree of Visual Reversal and Inversion Demonstrated by 5, 6, and 7 year olds."
R. Richards, Developmental Psychology (May 1971, vol. 31: 341-345).

2.2 "An Analysis of the Degree of Visual Reversal and Inversion Demonstrated by 5, 6, and 7 year olds."
research report by R. Richards.

3.1 "Factors Influencing Teachers' Engagement With The Teaching Profession."
S. Jackson, Instructor (Feb. 1971, 46: 27-31).

3.2 "Factors Influencing Teachers' Engagement With The Teaching Profession."
(S. Jackson, 1971)

4.1 Systems Study of Communication Patterns in Primary and Secondary Schools (with Emphasis on Urban Core Areas): Final Report.
T. Thomas, R&D Center Publication #18-311, December, 1970; available from ERIC, ED 013 988.

4.2 Final Report of Systems Study of Primary and Secondary Education (with emphasis on urban core areas) (ED 013 988)

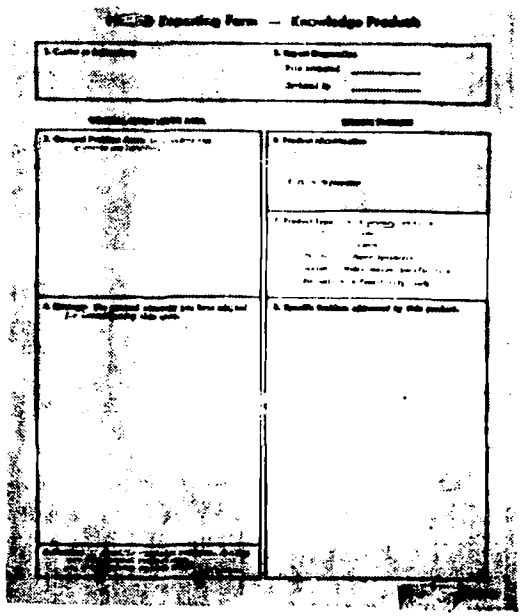
5.1 An Evaluation of the Athens Plan in Secondary Schools: Final Report.
J. Jones and H. Smith, Technical Report #472, XYZ Laboratory, Omaha, Nebraska, June 1971; also to be published in December 1971 under the title The Athens Plan by Easey Publishing Company (Lincoln, Nebraska).

5.2 An Evaluation of the Athens Plan in Secondary Schools: Final Report.
J. Jones and H. Smith, XYZ Laboratory, Omaha, Nebraska, June 1971.
The Athens Plan. (Inappropriately listed as a separate product)
Jones and H. Smith (Easey Publishing Company: Lincoln, Nebraska, December 1971).

6.1 "The Effects of Video-Feedback in Training Counseling Interns in Problem-Identification."
W. Barnes, Tape Recording of a speech made to National Association of Counselor Educators, November 1970; available from XYZ R&D Center, Catalog #84-796.

6.2 "The Effects of Video-Feedback in Training Counseling Interns in Problem Indentification."
W. Barnes, XYZ R&D Center, November, 1970.

SECTION 7. Product Type



Instructions

In general, there are five types of knowledge products:

(a) *Literature reviews*--literature summarized along lines not previously used (i.e., reviews of literature never before summarized or summarized along new conceptual lines, as in the Annual Review of Psychology or the Review of Educational Research).

(b) *Reports of research*--reports of studies designed to test educational hypotheses, investigate problems, or discover basic relationships;

(c) *Theoretical papers/syntheses*--analyses of existing research leading to the development of new insights, theories, or conceptualizations;

(d) *Systems or model designs/specifications*--designs and descriptions of the component parts, and interfaces among the parts of an educational system or a model for producing educational change; and

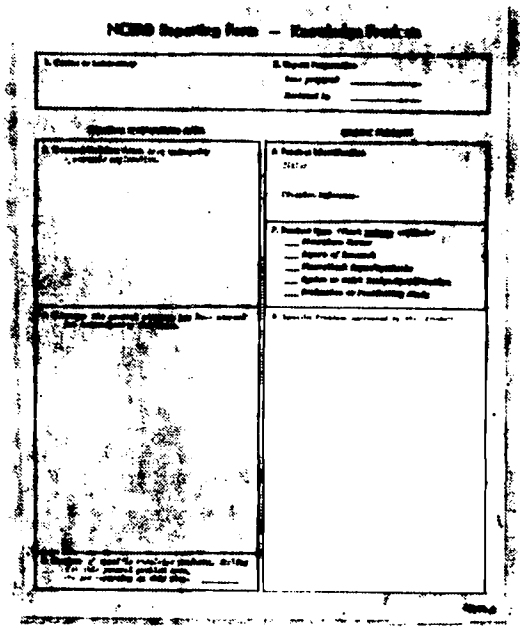
(e) *Evaluation or feasibility studies*--analyses of educational projects, or proposed projects, to assess their effectiveness, or feasibility.

If your product may be considered as consisting of two or more of these types, select that which is the primary emphasis of your product. For example, assume you have published a study on bilingual language patterns in three-year olds. As part of this study, you have: included a brief literature review relevant to this topic; conducted a research study testing your theories of bilingual language development; and in the discussion section, suggested a model for an early childhood education program based on your results. Although you have incorporated these three separate types of emphasis in one written document, you should classify the product in terms of its primary emphasis. In this case, you would probably wish to classify it as a research report testing your theories of bilingual language development. In the case of ongoing, programmatic efforts, report your most comprehensive treatment of your work to date. Clearly this latter would be a cumulative effort. The several independent research studies reported one year may be incorporated the following year into a synthesis of those results; and that, in turn, in a proposed new model for an instructional system or curriculum.

After you have decided the proper classification of your product, put an x on the appropriate line in Box 7.

Examples

1. Your agency surveyed a number of studies and papers about bilingualism and reading and from this information identified factors which seem to have the most harmful effect on reading performance.	This is an example of a "Literature Review."
2. Your agency administered the PQR test of visual perception to a large number of six-year-olds to determine how much six-year-olds, on the average, demonstrate visual reversal and inversion. A report describing this effort has been prepared.	This is an example of a "Report of Research."
3. The results of studies on why teachers enter and stay in the teaching profession were analyzed and a theory about which factors influence teachers' engagement with the teaching profession was proposed in the article.	This is an example of a "Theoretical Paper/Synthesis."
4. Your agency has been studying the communication patterns occurring within urban learning systems, identifying factors that affect the communication patterns, content of the communications, and motivational factors determining the content. You have prepared a model which describes the function of these three areas and the interface between the areas.	This is an example of a "System or Model Design/Specification."
5. Your agency has analyzed ongoing programs under the Athens Plan to ascertain the effects the Athens Plan has on its students.	This is an example of an "Evaluation or Feasibility Study."
6. Your agency is preparing a series of papers each of which outlines and evaluates alternative models and prepares recommendations for various aspects of a program of individualized education.	This is an example of a "System or Model Design/Specification."
7. You have been conducting a study to determine which of two formats, programmed instruction or seminar discussions, produces the greater amount of learning.	This is an example of a "Report of Research."
8. You have prepared a paper proposing variables that should be considered in designing administrative systems in secondary schools based on studies of various administrative systems across the country.	This is an example of a "Theoretical Paper/Synthesis."



SECTION 8. Specific Problem

Instructions

A great deal of specific information is normally required for the solution of even a modest educational problem. The problem you described in Section 3 is likely to be no exception. It will presumably require many new knowledge products. You will be preparing a product reporting form for each of the products you have generated.

In this section you should describe the specific problem that the product you are now reporting was intended to solve. In general, the statement of the specific problem is analogous to the problem statement of a journal article or doctoral thesis. It should describe as carefully and accurately as possible precisely what knowledge deficit your product was designed to rectify. To delineate the scope of the specific problem as clearly as possible, you should identify all of the parameters basic to the problem.

Note: This section should not contain a description of the goals of the product development effort, such as "to provide a set of guidelines for the effective planning and management of school districts;" nor is it intended to simply catalogue all of the hypotheses being tested.

Clear Examples

1.1 We need to identify what aspects of bilingualism seem to have had the most harmful effects on reading performance of primary grade children in previous research efforts to provide a starting point for our research efforts.

2.1 Information is needed about the degree to which visual inversion and reversal of geometric figures are present in primary grade minority group children. (This may be but one study in a long series subsumed under General Problems as stated in Section 3, Example 2.1.)

3.1 We need to identify elements of the classroom discipline situation which engage or alienate teachers from "discipline follow-through."

4.1 We need to develop a model for community leader-school official communication in urban learning systems by which the system can become more responsive to community needs.

5.1 Information is needed about the effects of the Athens Plan Rapid Feedback System on students' academic performance, as compared to the performance of control students.

Less Clear Examples

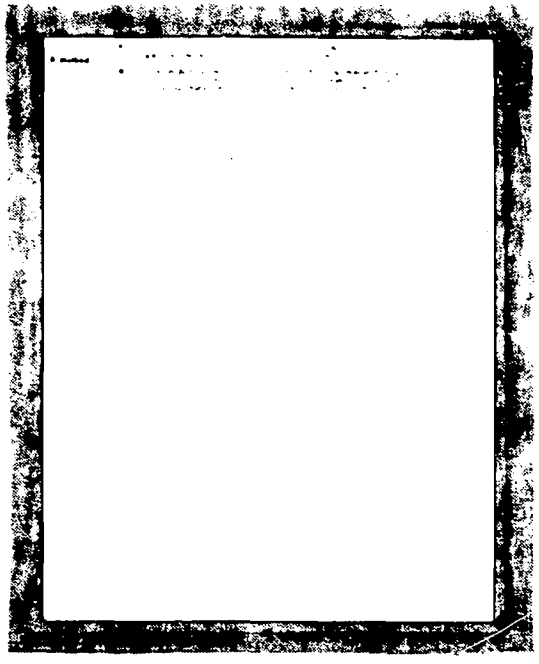
1.2 We don't know what the relationship is between bilingualism and reading.

2.2 We don't know how much visual inversion and reversal are present in children.

3.2 Although "good teachers" are committed to the teaching profession, little is known about the nature of that commitment.

4.2 As part of a new approach to the improvement of urban learning, we need a system analysis capability for the analysis of urban learning.

5.2 Our evaluation of the Athens Plan was prompted by a need for "hard facts" about its effectiveness.



SECTION 9. Method

Instructions

You should complete this section if your product is a literature review, research study, or evaluation and if your product does not indicate the method you used in its production. Methods will vary, depending on the type of product. For example, the procedures involved in conducting a literature review (identifying and retrieving references from primary and secondary sources) are quite different from those involved in designing a research or evaluation study (selecting a sample, identifying variables to be measured, locating appropriate assessment instruments, and selecting appropriate analytical techniques). Typically research reports include an extensive treatment of method. Literature reviews frequently do not. If your product does not indicate the procedures you used, please provide sufficiently detailed information so that reviewers familiar with the specific problem area may clearly understand the general procedures you followed.

Clear Examples

1.1 References to documents concerned with bilingualism and reading were obtained from ERIC, Psychology Abstracts, Education Index, and Dissertation Abstracts; a total of 673 references was obtained. Each reference was critically reviewed and abstracted. Each reference was then coded to indicate which of several specific kinds of information were treated in the document. (Kinds of information sought included characteristics of programs designed to improve the reading of bilingual children, effects of these programs, and variables of bilingualism associated with reading. The results of the review were then summarized and specific variables that seemed to have a negative effect on reading performance were identified from the results.

Less Clear Examples

1.2 Over 600 references were reviewed to find out what kinds of programs had been designed to improve reading of bilingual children, what the effects of these programs were, and what variables of bilingualism seemed to be associated with reading performance.

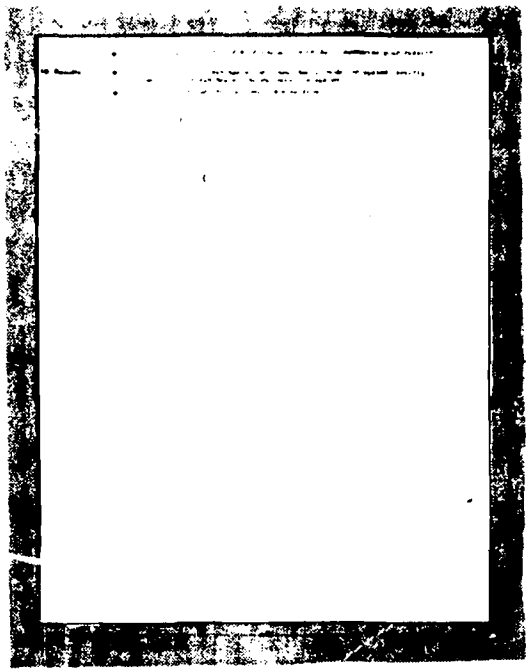
2.1 The first task was to identify an appropriate measure or measures of reversal and visual inversion. After surveying the available instruments it was decided to use the PQR test of visual perception. Next, a sample of 300 five-, six-, and seven-year olds, 30 from each of 10 schools in the state, was randomly selected. Each student was then administered form B of the PQR test of visual perception and the students' responses to the test items were recorded.

Six project staff members were involved in administering the test to students to control for experimenter/child interaction effects. The conditions of test administration were standardized. In addition, the following data were also recorded for each student: age in years and months, sex, and performance in reading activities (as measured by the ABZ Primary Achievement Test. Students' performance on items relating to visual inversion and reversal shift were then correlated with the other student variables using the BXY intercorrelation technique. (This is an example of a method used in conducting a research study.)

2.2 Three hundred students of age 5 to 7 were tested using the PQR test of visual perception. The data were analyzed together with information about the students' age, sex, and reading ability.

3.1 Our evaluation of the Athens Plan was based on data gathered about students presently participating in the Athens Plan, alumni, and parents. A sample of 150 Athens Plan students and alumni were randomly selected from the school attendance rosters and records. A group of 150 students from the same schools, but not enrolled in the Athens Plan, were selected to serve as controls. Basic data about the students' academic performance, such as GPA and class standing, were collected for all the students, both past and present. In addition, each student was asked to complete an open-ended questionnaire inquiring about his attitudes toward school; alumni were also asked about their post-high school experiences. A sub-sample of 20% of both students and alumni were interviewed, using the critical incident technique, about good and bad experiences they had in Athens Plan classes. Finally, questionnaires were sent to parents inquiring, again using the critical incident technique, about incidents in which they had had positive and negative contacts with the schools. These data were analyzed using Smithson's analysis of covariance. The results were incorporated in case history reports, or profiles, of the Athens Plan programs in the various schools.

3.2 Through the use of such evaluation techniques as survey instruments and personal interviews, we compiled descriptions of the Athens Plan as it is presently operating in nearby high schools.



SECTION 10. Results

Instructions

In this section you should summarize the results of your product. Specifically, you should tell exactly what new knowledge was generated. Depending on the nature of your product, your discussion will include different kinds of information.

If you have produced a research or evaluation study, you should interpret the data collected. In this case, your discussion would be similar to the results section of a journal article. You should specify both the areas of difference and the amount of difference that you discovered.

If you have produced a knowledge synthesis, theory, or model, then you should briefly describe the conclusions or model that resulted from your efforts. You might also discuss how the theory or model functions in practical applications, such as validation studies on pilot tests.

If your product was a literature review, you should omit this section.

In discussing the results of your product you should be as specific as possible. The examples below are, of course, only brief samples of the kind of information you should provide. You should use as much space as is necessary to adequately discuss the new knowledge obtained. This is one place where, contrary to opening instructions, you might wish to quote extensively from the abstract of your final report.

Clear Examples

1.1 Omit this section for a literature review.

2.1 Our studies showed a strong negative correlation (r=-.82) between children's age and the amount of visual inversion and reversal demonstrated (as measured by items 10-17 and 26-31 of the PQR Test of Visual Perception). There was a significant difference between the incidences observed for five- and six-year olds and those observed for seven year-olds ($p<.01$). Thus it seems that reversal and visual inversion tend to disappear in seven-year old lower SES children. In addition, a negative correlation

Less Clear Examples

2.2 Visual inversion and reversal both seem to be related to children's age and to their reading ability. There doesn't seem to be a relationship between these variables and the children's sex.

was found between incidence of visual inversion and reversal and student performance in reading (as measured by the ABZ Primary Achievement test). No correlation was found between incidence of reversal and visual inversion and the sex of the child. (Results of a research study.)

3.1 From our study of research already done in this area, we have theorized that the single most important element contributing to a teacher's engagement with or commitment to, the teaching profession is student characteristics. Of particular interest to us was the evidence suggesting that elements commonly considered to be of importance (teacher interaction with colleagues and role in decisionmaking) appeared to contribute little to teacher engagement. (Results of a knowledge synthesis.)

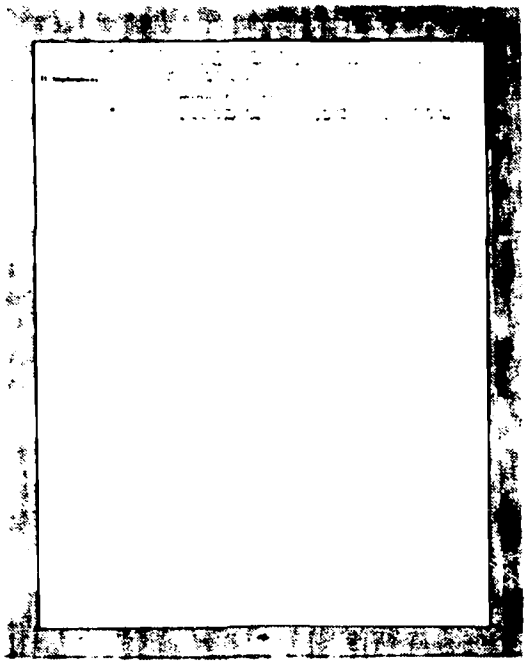
3.2 Among the most important factors in influencing teacher commitment to the teaching profession is student interest in academic subjects. Of lesser importance are such factors as teacher participation in decision making and teacher interaction with colleagues.

4.1 The model developed by our staff focused on transactions (communications conveyed by verbal and nonverbal modes) at three levels of detail with three corresponding submodels. The first level is concerned with the broad groups of transactions that occur in urban learning systems and the factors that affect them, e.g. demographic trends or financial resources. The next level of detail is a submodel for analyzing the content of those transactions: the content is viewed in the context of the demands and resources of the participants. Finally, the last level of detail is developed in a submodel that deals with the motivational factors that determine transactional content. Attachment A provides a graphic illustration of this model. (Results of a model design.)

4.2 See Attachment A.

5.1 Students in the Athens Plan perform significantly better than control students on measures of academic achievement (differences in mean GPA, $p<.05$). In addition, AP students expressed more satisfaction with their classes, reporting a significantly greater number of incidents of positive experiences. However, no significant difference was discovered between the responses of parents of AP students and other parents.

5.2 There are several schools in which the Athens Plan, by several criteria of effectiveness, is operating effectively and successfully. More Athens Plan students think their high school education was useful preparation for further schooling and employment, and Athens Plan students tend to do better in their school work.



SECTION 11. Implications

Instructions

In general, the implications you draw concerning your knowledge product will consist of its potential contributions to education, both those that could conceivably occur and those that are likely to occur. You should first discuss these implications in terms of the audience for whom your product was intended. That group would be defined as those educational professionals who will be the immediate users of your product.

Second, you should discuss the implications of your product for education in general. In other words, what might be the consequences of your product for educational programs, schools and school systems, teachers and teacher training, and so forth.

Finally, if it is appropriate, you should discuss the implications of your product for children or students. What changes might occur in children or in experiences provided to children as a result of your product.

Remember, you should discuss both the possible, or conceivable, consequences and also the probable, or likely, consequences of your product.

If your product itself goes into detail regarding its possible and probable implications, you may wish to only briefly summarize the implications here and cite the appropriate section/pages of your product. Be sure your product discussion does in fact treat immediate practical implications as well as the longer-range hypothetical implications.

Clear Examples

1.1 The literature review enabled us to identify key variables and develop hypotheses for our subsequent research studies on the effects of bilingualism on reading performance. It would also provide similar guidelines for research others wished to do in this area. The findings of this review may also influence the types of reading programs developed for bilingual children. Finally, bilingual children may learn to read better as a result of this effort.

Less Clear Examples

1.2 Our review may affect the kind of work done by other researchers and developers and, ultimately, the reading of bilingual children.

2.1 One potential consequence of this study would be that all reading programs designed for five and six year olds would be discarded since children do not appear to have resolved reversal and visual inversion problems until age seven. This would, in turn, save school districts the money spent on these programs, allow teachers to delay the introduction of reading until second grade, and eliminate the sense of failure engendered in students by trying to teach them to read too soon. What is more likely to occur are studies designed to identify methods of fostering the development of children's skills in visual inversion and reversal resolution.	2.2 This study could change the programs offered to five and six year olds.
3.1 It is likely that teacher educators will attempt to promote the teachers' understanding of his own professional commitment, based on the results of our study. It is also possible that this information might be considered in admitting candidates to teacher training programs.	3.2 Teachers should be made more aware of their reactions to student characteristics. In particular, we feel that designers of teacher training programs should take our results into consideration in developing their future training programs.
4.1 Our model could result in increased communication between community leaders and school officials; this, in turn, could make the school more responsive to the needs of the community and also increase the involvement of community agencies in the schools. An additional effect would be that students would see their schooling as more relevant to the "outside" world.	4.2 From our preliminary study, we feel that our test of a systems analysis model indicated that systems analysis can significantly augment a state or local government's policy-making capability. Thus, we would recommend that state and local governments responsible for school planning should consider implementing systems analysis methods in the next few years.
5.1 The effectiveness of the Athens Plan, as demonstrated in this study, could result in the program being adopted on a wide scale. This would probably result in improved academic performances of the students. It could also have implications on teacher training programs in that teachers would have to be trained more for individual and small group instruction. Because of the cost of adopting the entire program, it is more likely that schools will adopt or adapt specific components of the program. This would have implications for the researchers who would be involved in adapting the program for specific schools. Also, teaching behavior and student academic performance would be affected to the extent of the schools' involvement in the Athens Plan.	5.2 Analysis of the effectiveness of the Athens Plan suggests that it should be considered for wide scale implementation.

THANK YOU FOR YOUR EFFORT IN PRODUCT REPORTING.

All reports of products should be sent to your agency director, or his designee, for review, approval, and signature in Section 3, prior to transmittal to NIE.

Product Reporting Form — Knowledge Products

1. Center or Laboratory

2. Report Preparation

Date prepared: ____________

[illegible] ____________

GENERAL KNOWLEDGE AREA

3. General Problem Area: *Area undergoing systematic exploration.*

4. Strategy: *The general strategy you have adopted for investigating this area.*

5. Number *of specific knowledge products, dealing with this general problem area, you are reporting at this time.* ________

SPECIFIC PRODUCT

6. Product Identification

Title:

Citation reference:

7. Product Type *(Check primary emphasis)*

- ___ *Literature Review*
- ___ *Report of Research*
- ___ *Theoretical Paper Synthesis*
- ___ *[illegible] Design Specification*
- ___ *Evaluation or Feasibility Study*

8. Specific Problem *addressed by this product.*

10-71-B (D)

9. Method

- *If specific product is research, literature review, or evaluation, describe the method you used (or will use) in detail. This section may be omitted if the specific product adequately describes the method used in the production of your results.*

- *If your product is an analytical paper synthesizing research results, or a specifications paper dealing with the parameters and/or operating characteristics of a new model or system, omit this section.*

10. Results

- *If your product is a research or evaluation product, briefly summarize your results.*
- *If it is a knowledge synthesis, or a new theory, model or system, briefly summarize your synthesis, theory, model or system.*
- *If it is a literature review, omit this section.*

11. Implications

- *Discuss this implications of your product:*

  1.) For your intended audience

  2.) For education in general; and

  3.) If appropriate, for children.

- *Discuss not only the theoretical (i.e. conceivable) implications of your product but also the more <u>probable</u> implications of your product, especially over the next decade.*

Appendix B

EVALUATORS MANUAL AND PRODUCT RATING FORMS

EVALUATORS MANUAL

PRODUCT RATING FORMS

EVALUATORS MANUAL

FOR THE

EVALUATION OF FEDERALLY SUPPORTED

RESEARCH AND DEVELOPMENT PRODUCTS

1972

This manual has been prepared by the
American Institutes for Research under
a United States Office of Education
Contract, Number OEC-0-70-4891.

TABLE OF CONTENTS

INTRODUCTION

The evaluation procedure described herein is the result of a joint effort by NCERD and OPPE to develop a procedure for the evaluation of products produced by Regional Educational Laboratories and university-based Research and Development Centers.

One important characteristic of this effort has been the involvement of laboratory and center directors, to the extent they found it possible, in the development, review, and critique of the procedures and criteria constituting this evaluation system.

The purpose of this manual is to orient the evaluator to his task, to describe the steps of the evaluation process, to give the evaluator general instructions with regard to his task, to describe in detail the criteria he is to use in carrying out his evaluation, and to summarize some background information about the history of the laboratory and center movement.

The procedures described have been developed to provide for the impartial evaluation of two types of laboratory and center outputs: knowledge products and the so-called "hard" developmental products.

No claim is made for the appropriateness of these procedures for the evaluation of other socially significant contributions made by laboratories and centers. These procedures do not provide for the evaluation of: community service; the development of an institutional capability to engage in educational R & D; manpower training contributions; or the like. Neither are these procedures appropriate for evaluating the management of laboratories and centers.

The Evaluation Paradigm

The evaluation of laboratory and center products may be described in terms of a series of steps. These sequential activities are as follows:

Phase I - Product Reporting

1. Product Reporting Forms and instructions for their completion are sent to Laboratories and Centers.

2. Laboratory and Center staff complete the Product Reporting Forms. The forms are reviewed by Laboratory and Center Directors before they are released.

3. The Product Reporting Forms are received by the evaluation coordinator, reviewed for completeness of information, and classified by product type.

Phase II - Product Evaluation

4. Letters are sent to the Laboratory and Center Directors notifying them of the pending evaluation summarizing the evaluation procedure, and request nominations for evaluations. Nominations are also requested at this time from other sources as well.

5. The evaluation coordinator submits the resultant list of nominees for each topic to the Laboratory and Center Directors and to the Office of Education for approval.

6. Laboratory and Center Directors are notified of the products selected for evaluation, copies of the products are requested, and confirmation of the information on the Corresponding Product Reporting Form is sought.

7. The evaluation coordinator selects six to nine evaluators for each product type from the list of approved

evaluators. Each panel will consist of at least three specialists in the topic area, one research and development specialist, and one consumer representative. Evaluation panel members will generally serve for two years.

8. Evaluators meet for an orientation-training conference. During this meeting evaluators are oriented to the evaluation procedure, review the criteria to be used, and execute several practice evaluations. After that, products not convenient for mail distribution are evaluated. Products amenable to mail distribution, or which require special field visits, will be evaluated subsequently.

9. Evaluators review products and make their initial evaluations independently. Completed Evaluation Forms are submitted to the evaluation coordinator, who will then circulate them within the panel prior to asking panelists to confirm their judgments. This step is intended as information exchange among the panel members so that they may, if they wish, reconsider their initial evaluation. Panelists' names will not be associated with their judgments during this information exchange process.

GENERAL INSTRUCTIONS

From the evaluator's point of view, after the orientation and training conference, the first step in evaluating a product is to review the Product Reporting Form which accompanies it. This form serve two purposes: one, it provides the evaluator with an overview of the product; two, it provides the evaluator with information which is often not available elsewhere.

In the first instance, the Product Reporting Form serves as a sort of product guide. Such a guide is particularly helpful for those products which are somewhat complex. The form provides a brief resume of the origin of the product, the number of "pieces" to the product, and the like.

The product report also indicates the level of development of the product. Depending upon the nature of various cooperative publishing arrangements that may have been established, it is quite possible that some products may be considered "completed" by their developing agency, and hence submitted for evaluation, even though the product is not "market-ready." Should this be the case, allowance must be made by the evaluator when he evaluates the product.

Information on the Product Reporting Form should be taken quite literally. These forms have been carefully prepared by the appropriate laboratory or center staff, reviewed by the director of the agency, and confirmed again by the director just prior to product evaluation.

If, in the course of product evaluation, an evaluator feels he would like additional information of some type regarding the product, he should request this information of the evaluation coordinator who,

when he obtains it, will communicate that information to all individuals evaluating the product in question.

After becoming familiar with the product, through the Product Reporting Form, the evaluator should then thoroughly inspect the product itself. The goal here should be one of maximum thoroughness.

After the product has been reviewed, the evaluator should next review any support documents that the product developer has submitted to substantiate claims about his product. If any such documents have been submitted, they accompany the product.

Finally, the evaluator should review the appropriate criterion list in this manual and then evaluate the product using the appropriate Product Evaluation Form.

The Product Evaluation Form for knowledge (research) products differs from that for developmental products. Therefore, the evaluator should verify the form he is using. He should also verify the product number on the top of the Evaluation Form against the product number on the top of the Product Reporting Form, and verify the product being evaluated against the product as it is described in the Reporting Form.

Upon completion of the task the evaluator should return the completed forms to the evaluation coordinator and complete any other supplementary instructions that might have been given.

CRITERIA - DEVELOPMENTAL PRODUCTS

A. Importance of General Problem

A problem is a recognized discrepancy between an existing state in education and a desired end state. As such, it may be described as an "educational need." In considering the importance of a problem, the question is "how crucial is it?" The magnitude of importance is a function of the number of people it affects and the intensity with which it affects them. A problem which intensely affects a large number of people is, of course, easily recognizable as an important problem. A problem that affects relatively few people, and only slightly, is easily recognized as being of little importance.

The difficulty of judging the magnitude of a problem's importance comes when judgments have to be made with regard to products affecting only a few persons but relatively intensely, as in the case of some special education programs. Difficulties may also be encountered with products that affect a larger number of people, but only modestly. It is at this point the judgment of a problem's importance must be tempered by one's philosophy, experience, and professional commitment.

B. Relevance of Product to General Problem

Relevance refers to the degree to which the product under consideration clearly and directly relates to the stated educational problem. The product that is addressed directly to the heart of the problem has greater relevance than the product which deals only with some tangential aspect of the problem. For example, if the product developer indicates that his product is intended to help solve the problem of chronic

poor reading in minority group children, a teacher's manual enhancing the story-telling abilities of primary grade pupils would be judged less relevant to the problem than a manual telling the teacher how to manipulate reinforcement techniques during reading instruction. This is not to say that the former product is not related to the teaching of reading; indeed, there are many who feel that verbal language ability is a necessary prerequisite to the enhancement of reading achievement: the product simply is not central to the problem as it was stated.

C. Comprehensiveness of the Product as Problem Solution

The comprehensiveness of a product depends on the degree to which the product meets the entire problem. If a product addresses all of the major facets of a problem, no matter how small or trivial the problem, then the product should be judged comprehensive. On the other hand, a product which deals with only a small portion of the general problem must be viewed as less comprehensive, regardless of the size of the effort devoted to the development of the product. It is not the size of the problem addressed which defines comprehensiveness; nor is it the size of the effort undertaken in the development of the product that counts. Rather, the extent to which the product addresses the whole problem, as stated on the product report form, serves as the basis for the evaluation on this criterion.

D. Content Accuracy

Accuracy refers to the extent to which facts, calculations, data, concepts, etc. presented in the product are informationally correct.

E. Content Clarity

Clarity refers to the extent to which the text or materials are clear in their message. The materials should be easily read and understood. Directions for their use should be offered in a straightforward manner. The user, whether he be student, teacher, administrator, etc., should not have to spend inordinate amounts of time trying to comprehend what is in the materials, the purpose of their existence, or how to use them.

F. Effectiveness

A product is effective to the extent that it works, i.e., to the extent that it meets its intended objectives.

The product per se typically does not include information on its effectiveness. The evaluator normally must base his judgment of the product's effectiveness on an examination of the reports and support documents submitted by the developing agency. A brief discussion of effectiveness may be found in Section II, Product Outcomes, on the Developmental Product Reporting Form. Support documents, if any, accompany the product.

If an evaluator has information or knowledge about the effectiveness of the product under consideration, from sources other than those documents submitted in support of the product by the developing agency, that evaluator should so notify the evaluation coordinator so that the additional evidence may also be made available to the other evaluators. In other words, evaluators should be careful to avoid judging the effectiveness of a product on the basis of either opinion or prior judgment made as a consequence of evaluation results not currently supplied with the product, and thus, not available to other evaluators. The judgment of product effectiveness must be

based on a careful review of objective data.

Of course, if the product developer does not supply any evidence in support of his product's effectiveness, no judgment of product effectiveness can be made. The lack of any supporting evidence should be so indicated on the product evaluation form.

G. Reasonable Cost to Adopt/Implement Given Outcome

This criterion applies to what is commonly referred to as "purchase price." The question here is whether the product is worth purchasing given what it is expected to do. In some cases this question is fairly easy to answer. For example, a program which improves children's knowledge of classical composers for $20 per pupil per year would probably be judged as relatively expensive. On the other hand, some comparable expenditure, or even a considerably higher one, may be happily accepted if the outcome of the expenditure is highly valued. For example, it might cost many thousands of dollars to institute a new reading program. However, if it were effective in raising the average reading level of non-readers to a level of independent reading competency, it would be judged well worth the cost.

The main question here is not whether the cost of adoption is high or low, but whether the cost is reasonable given what the product will do, i.e., whether the educational community is likely to get a good return for its investment.

H. Reasonable Cost to Use/Operate Given Outcome

This criterion is related to what is often called "operating costs." It applies to such routine ongoing expenses as replacement

of consumable materials, equipment repair and servicing, periodic personnel costs, and the like. These are costs necessary for the continued use of a product after it has been acquired and installed. The question here is once again not whether the costs for continued operation of the product are high or low, but rather, whether the expenditure of funds for continued operation is worthwhile, given the results accruing from product use.

I. Potential Market

Potential market refers to the number of possible clients for the product. Here the emphasis is on the possible market for a product dealing with this problem, not on the probable sales for this particular product. That is, what would be the potential size of the market if the product were effective and attractive, and clients could afford its purchase?

While it is recognized that a number of qualifiers affect the realistic boundaries of potential markets, evaluators should nonetheless attempt to make a judgment about the possible scope of utilization of a product. Some products, while very important, may be pertinent for only limited audiences. Thus, such products would have quite a limited potential market. Other products might have more general or pervasive application throughout all educational audiences. Products which contribute to solutions of more pervasive problems would have a wider potential market.

J. Potential Marketability

The question here is "Do you think the product, as it is presently formed, will lend itself to effective marketing?" That is, will someone be able to market it effectively? A number of factors enter into

this decision: Is the product attractive? Is it assembled in such a way that it can be efficiently produced? Does it lend itself to convenient advertising, supply, classroom storage, etc.?

K. Potential Impact

In assessing potential impact, evaluators should ask to what extent the product has the potential for improving educational practice on a major scale. The basic question is to what extent the product is likely to effect a change in educational practice considering all the characteristics of the product and other factors which may influence its adoption and utilization.

CRITERIA - KNOWLEDGE PRODUCTS

A. Importance of General Problem

A problem is a recognized discrepancy between an existing state in education and a desired end state. As such, it may be described as an "educational need." In considering the importance of a problem, the question is "how crucial is it?". The magnitude of importance is a function of the number of people it affects and the intensity with which it affects them. A problem which intensely affects a large number of people is, of course, easily recognizable as an important problem. A problem that affects relatively few people, and only slightly, is easily recognized as being of little importance.

The difficulty of judging the magnitude of a problem's importance comes when judgments have to be made with regard to products affecting only a few persons, but relatively intensely, as in the case of some special education programs. Difficulties may also be encountered with products that affect a larger number of people, but only modestly. It is at this point the judgment of a problem's importance must be tempered by one's philosophy, experience, and professional commitment.

B. Relevance of Product to General Problem

Relevance refers to the degree to which the product under consideration clearly and directly relates to the stated educational problem. The product that is addressed directly to the heart of

the problem has greater relevance than the product which deals only with some tangential aspect of the problem. For example, if the product developer indicates that his product is intended to help solve the problem of chronic poor reading in minority group children, research on the story-telling abilities of primary grade pupils would be judged less relevant to the problem than research on how to manipulate reinforcement techniques during reading instruction. This is not to say that the former product is not related to the teaching of reading; indeed, there are many who feel that verbal language ability is a necessary pre-requisite to the enhancement of reading achievement: it simply is not central to the problem as it was stated.

C. Comprehensiveness of the Product as Problem Solution

The comprehensiveness of a product depends on the degree to which the product meets the entire problem. If a product addresses all of the major facets of a problem, no matter how small or trivial the problem, then the product should be judged comprehensive. On the other hand, a product which deals with only a small portion of the general problem must be viewed as less comprehensive, regardless of the size of the effort devoted to the development of the product. It is not the size of the problem addressed which defines comprehensiveness; nor is it the size of the effort undertaken in the development of the product that counts. Rather, the extent to which the product addresses the whole problem, as stated on the product report form, serves as the basis for the evaluation on this criterion.

D. Originality of Product

An original product is one which represents an imaginative or ingenious approach to solving the general problem to which the product is addressed.

The originality may be in problem conceptualization, methodology, or interpretation. The uniqueness of the document's ideas and/or methodology, of course, may only be judged within the evaluator's knowledge and experience.

E. Quality of Literature Discussion

It is clear that for most types of knowledge products, customary literature reviews provide a strong integrating context. The desirability for comprehensiveness varies with the type of knowledge product. Products whose sole purpose is to review the literature need be, of course, very comprehensive. Citations should include all the major efforts in an area and probably many of the lesser known efforts. However, for most types of knowledge products, the review may be less than comprehensive in the usual sense, but it should be directly related to the specific problem addresses in the documents. In all cases, the review should: a) be appropriate to the specific problem area; b) make explicit the relationship of previous research to the problem area cited; and c) point out how the additional new research accommodates or enhances the previous citations. In addition, the researcher should exhibit: a) an appreciation of the current "state of the art"; b) total familiarity with recent, pertinent literature; and c) an attempt to interpret, synthesize, and evaluate the relevant literature.

F. Adequacy of Research Design

Obviously, like originality, the criterion of design adequacy includes a variety of considerations. Clearly all conceivable aspects of design cannot be evaluated at this time. This evaluation, thus, must be somewhat "holistic."

Not all types of knowledge products will include a formal research design as an integral aspect of the presentation. A discussion of design is not likely to be included in literature reviews, for example. However, it is very likely to be a part of reports of research and evaluation or feasibility studies.

If it is present, basic consideration should include:

a) the degree to which the design is suited to the problem;
b) whether the design represents a rigorous test of the stated or implied hypotheses;
c) whether careful attention has been directed toward reducing sources of error and minimizing threats to validity such as:
    1) random assignment of subjects,
    2) statistical or experimental control of intervening variables,
    3) sufficient numbers of subjects,
    4) dependent variable instruments of sufficient validity and reliability,
    5) sampling which allows for justifiable generalizing, or
    6) acknowledgment and satisfaction of statistical assumptions.

Since a number of factors will be under consideration in this criterion, evaluators may wish to make explanatory notations of their ratings in the Comments section.

G. Appropriateness of Interpretation

Appropriateness of interpretation, deals with the degree of reasonable accord between the factual results of a study and the statements made about those results. The key issue is the degree to which interpretations or statements about the results are, in fact, justified by the data. Evaluators should be alert to misinterpretations, inappropriate generalizations, and the like.

H. Reasonableness of Conclusions/Recommendations

This criterion relates to judgments about those statements which go beyond simple interpretation of results. The consideration here is the degree to which a researcher is justified in "making something" of his findings. The evaluator should be alert to the "tightness" of these statements; that is, do they follow the general design? Are his conclusions substantiated? exaggerated? modest? Has he gone beyond his data? In general, the main issue is whether the discussion or the conclusions are related to the design, substantiated by the data, and generally logical.

I. Clarity of Presentation

For the most part, this criterion speaks for itself. The key consideration is the degree to which the effort has been logically organized and described in plain, straightforward language making it easy to follow and understand. The problems, concepts, hypotheses, conclusions, and so forth should be clearly and logically stated. In addition, the project should be so described as to make it completely comprehensible and, in appropriate types of research, replicatable.

J. Potential Impact

In assessing potential impact, evaluators should ask to what extent the product has the potential for improving educational practice on a major scale. The basic question is to what extent the product is likely to effect a change in educational practice, considering all the characteristics of the product and other factors which may influence the adoption and utilization of its concepts.

BACKGROUND INFORMATION

In 1963, the Research and Development Centers Program was established under the then-existing provisions of the Cooperative Research Act. Between 1964 and 1967, ten research and development centers were established at major universities across the country. Their mission was to conduct basic and applied research and exploratory development in designated educational areas through large-scale, cooperative efforts.

In 1965, additional legislation was passed providing for the establishment of a series of independent, non-profit, educational development corporations. These were called Regional Educational Laboratories. Their mission, like the university-based R & D centers, was to engage in educational research and development within specific geographical regions. Twenty laboratories were established in 1966.

All told, a total of thirty laboratories and centers were established by USOE. In addition, two research and development centers focusing on vocational education were established during this period. As of Spring 1972, nine laboratories and two R & D centers have been discontinued, leaving a total of eleven Regional Laboratories, eight Research and Development Centers, and two Vocational Research Centers.

Through 1969, a total of approximately $114 million had been spent on the original thirty-two agencies. In FY '70 and '71, an additional $44 million were awarded the eleven remaining Regional Laboratories and $15.5 million were granted the eight remaining R & D Centers.

During this same period, another $5.3 million went to four Regional Laboratories no longer operating as of Spring, 1972. Therefore, since their inceptions, Laboratory and Center funding has totaled more than $180 million.

Excluding the two vocational centers, the now-operating eight R & D Centers and eleven Laboratories represent a total investment of $141 million through FY '71.

Annual funding of laboratories and centers has ranged from $500,000 to $3.5 million per year. Briefly speaking, laboratories and centers may be divided into three funding groups: (a) those funded most heavily, on the order of $3 to $4 million per year; (b) those funded with intermediate funding, i.e., on the order of $2 to $3 million a year; and (c) those with funding of approximately $500,000 to $1.5 million per year. The various laboratories and centers may be roughly classified as follows:

Group A

Research for Better Schools, Inc.
Southwest Regional Laboratory

Group B

Far West Laboratory
Central Midwestern Regional Educational Laboratory
Southwest Educational Development Laboratory
Center for Urban Education
Northwest Regional Educational Laboratory
Learning Research and Development Center
Center for R & D for Cognitive Learning

Group C

Appalachia Educational Laboratory
Stanford Center for R & D in Teaching
Southwestern Cooperative Educational Laboratory
National Laboratory for Higher Education
Mid-Continent Regional Educational Laboratory
Center for R & D in Higher Education
Center for the Study of Evaluation
Center for the Advanced Study of Educational Administration
Center for Social Organization of Schools

For reference purposes, the names and locations of the twenty-one remaining laboratories and centers are as follows:

Regional Educational Laboratories

Appalachia Educational Laboratory (AEL)
Charleston, West Virginia

Center for Urban Education (CUE)
New York, New York

Central Midwestern Regional Educational Laboratory (CEMREL)
St. Ann, Missouri

Education Development Center, Inc. (EDC)
Newton, Massachusetts

Far West Laboratory for Educational Research and Development (FWLERD)
Berkeley, California

Mid-Continent Regional Educational Laboratory (McREL)
Kansas City, Missouri

National Laboratory for Higher Education (NLHE)
Durham, North Carolina

Northwest Regional Educational Laboratory (NWREL)
Portland, Oregon

Research for Better Schools, Inc. (RBS)
Philadelphia, Pennsylvania

Southwest Educational Development Laboratory (SEDL)
Austin, Texas

Southwestern Cooperative Educational Laboratory (SWCEL)
Albuquerque, New Mexico

Southwest Regional Laboratory for Educational Research and Development (SWRL)
Inglewood, California

Educational Research and Development Centers

Center for Research and Development for Cognitive Learning
University of Wisconsin

Center for the Advanced Study of Educational Administration
University of Oregon

Center for Research and Development in Higher Education
University of California at Berkeley

Research and Development Center in Teacher Education
University of Texas

Learning Research and Development Center
University of Pittsburgh

Stanford Center for Research and Development in Teaching
Stanford University

Center for the Study of the Evaluation of Instructional Programs
University of California at Los Angeles

Center for the Study of the Social Organization of Schools and
the Learning Process
Johns Hopkins University

Vocational Centers

Center for Occupational Education
Raleigh, North Carolina

Center for Vocational and Technical Education
Columbus, Ohio

PRODUCT RATING FORMS

Date_____

Product Number_____

Evaluator_____

DEVELOPMENTAL PRODUCT RATING FORM

The following are abbreviated definitions of the criteria used to evaluate developmental products. More elaborate definitions are offered in the Evaluators' Manual.

A.	IMPORTANCE OF GENERAL PROBLEM:	. . . degree to which problem is crucial to education . . . magnitude of the problem
B.	RELEVANCE OF PRODUCT TO GENERAL PROBLEM:	. . . degree to which product clearly and directly relates to stated problem
C.	COMPREHENSIVENESS OF THE PRODUCT AS PROBLEM SOLUTION:	. . . degree to which product meets the whole problem
D.	CONTENT ACCURACY:	. . . informationally correct . . . a precise accounting and presentation
E.	CONTENT CLARITY:	. . . an easily understood exposition . . . full, unambiguous explanations and directions
F.	EFFECTIVENESS:	. . . degree to which product solves the problem . . . degree to which product meets its objectives
G.	REASONABLE COST TO ADOPT/ IMPLEMENT, GIVEN OUTCOME:	. . . degree to which product is worth buying, given what might or will come of its use
H.	REASONABLE COST TO USE/ OPERATE, GIVEN OUTCOME:	. . . degree to which product is worth continuing to use
I.	SCOPE OF POSSIBLE MARKET:	. . . possible number of users, buyers, clients
J.	AMENABILITY TO MARKETING:	. . . attractiveness of product . . . ease of acquisition and use
K.	POTENTIAL IMPACT:	. . . likelihood of effecting change in educational practices, given all factors

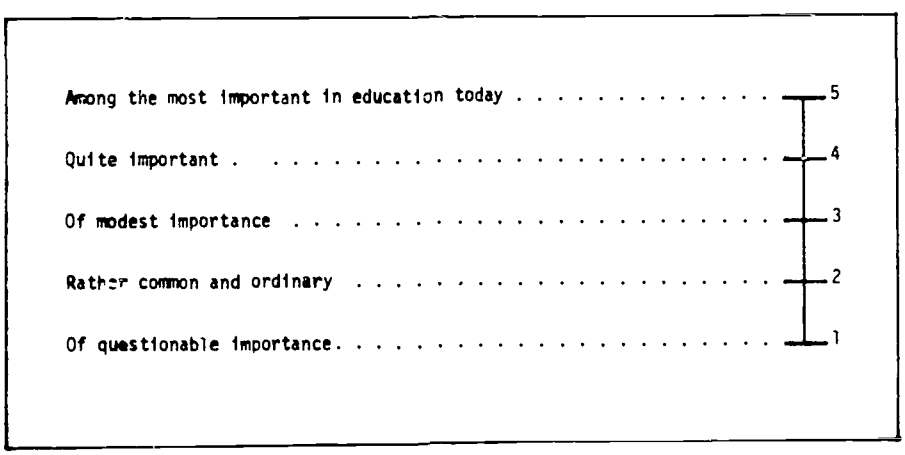
4-72/ID

INSTRUCTIONS

For each scale, select that phrase which best represents your judgment of the product. Then circle the number of that phrase. Do not mark intermediate points.

Should you, for some reason, be unable to arrive at a rating on a particular criterion, note this and explain why in the Comments section. Also use the Comments sections for any additional remarks you may wish to make. Comments explaining very low ratings will be especially helpful. For the final criterion, Potential Impact, please explain why you feel the product will or will not have impact on the educational community.

A. PROBLEM IMPORTANCE

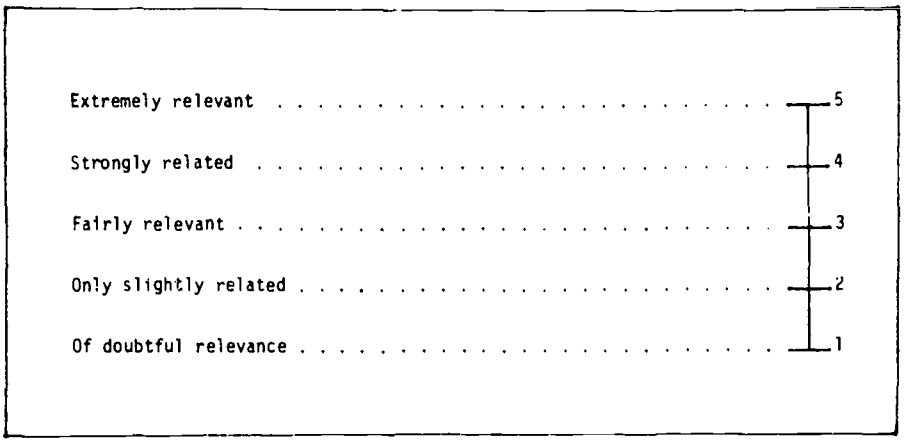


Comments: _____

_____

_____

B. RELEVANCE OF PRODUCT TO GENERAL PROBLEM

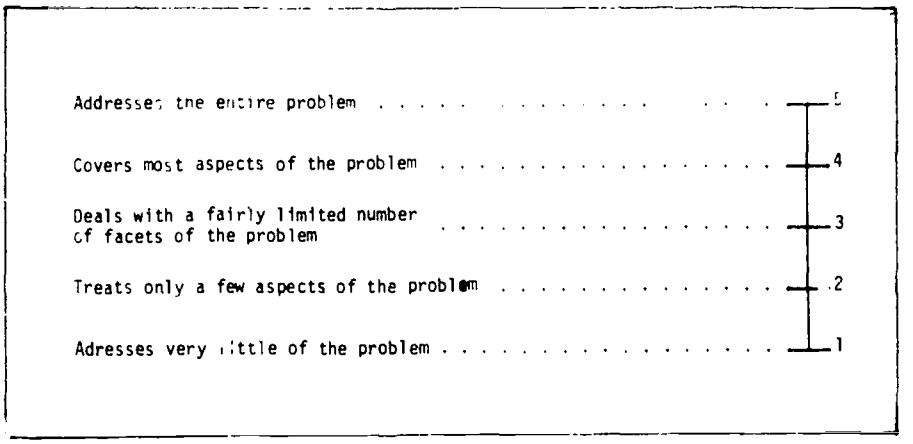


Comments: _____

_____

_____

C. COMPREHENSIVENESS OF PRODUCT AS PROBLEM SOLUTION

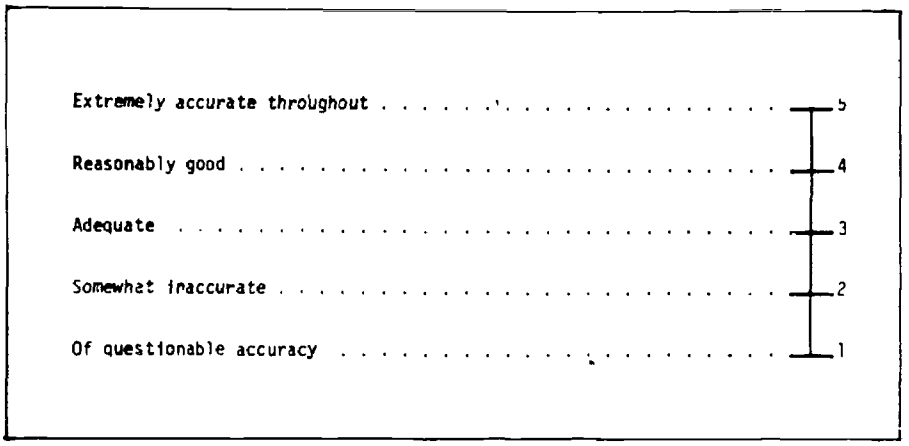


Comments: _____

_____

_____

D. CONTENT ACCURACY

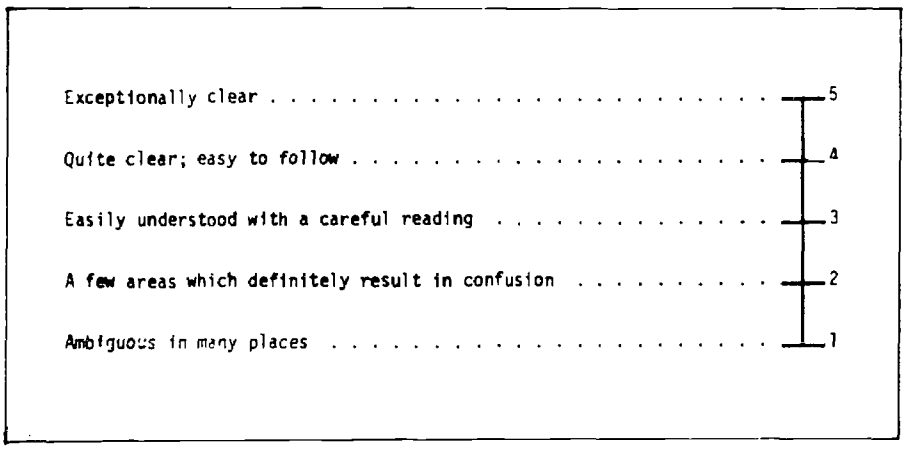


Comments: _____

_____

_____

E. CONTENT CLARITY



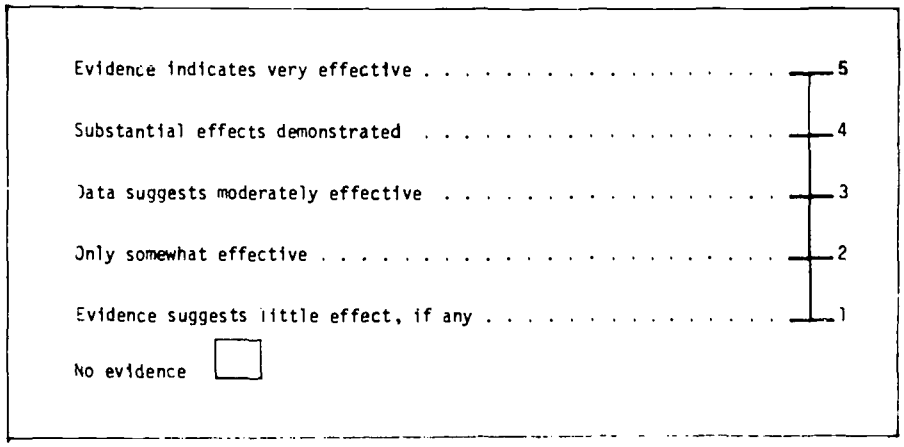
Comments: _____

_____

_____

F. EFFECTIVENESS

Note: if there is no evidence on which to judge the effectiveness of this product, indicate by checking the box labeled "No Evidence."

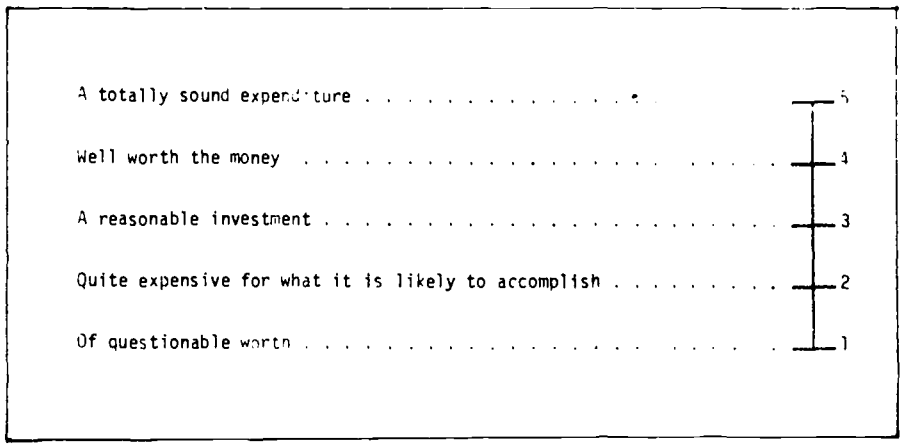


Comments: _____

_____

_____

G. REASONABLE COST TO ADOPT/IMPLEMENT, GIVEN OUTCOME

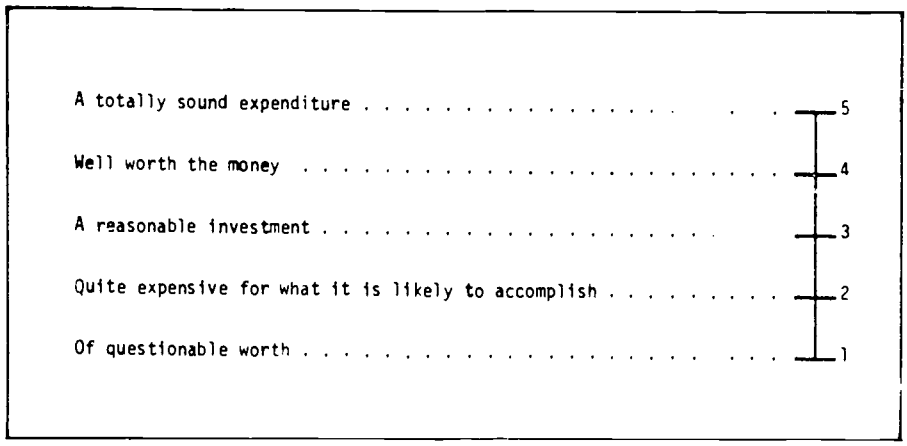


Comments: _____

_____

_____

H. REASONABLE COST TO USE/OPERATE, GIVEN OUTCOME

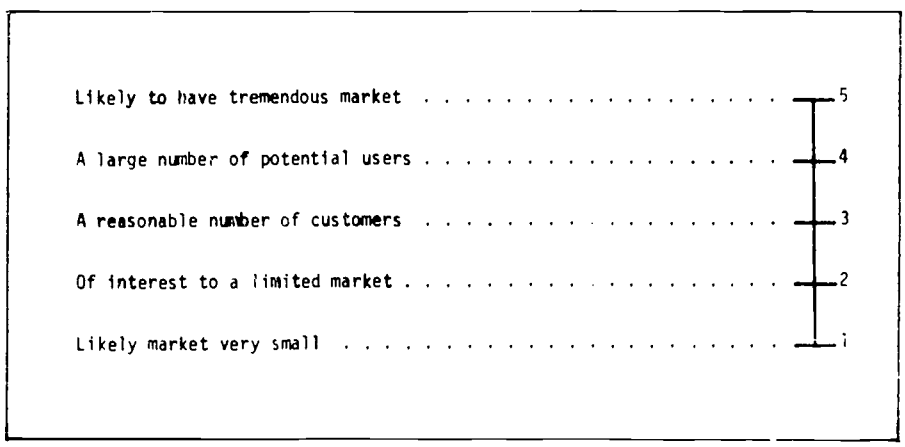


Comments: _____

_____

_____

I. POTENTIAL MARKET

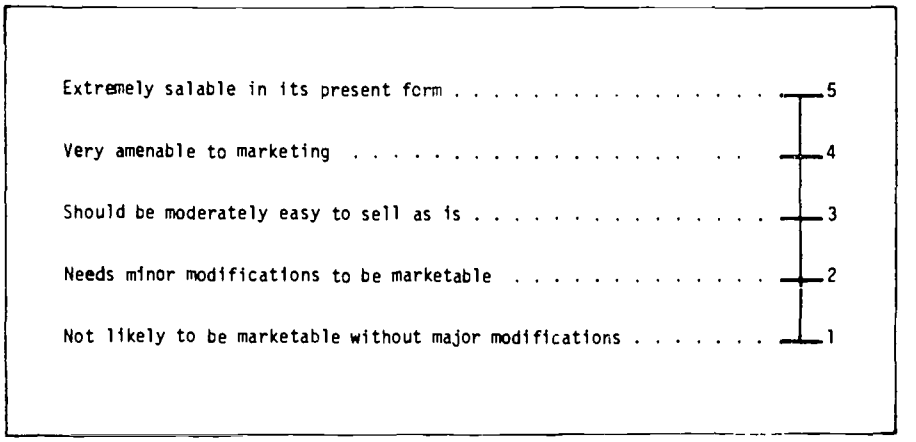


Comments: _____

_____

_____

J. POTENTIAL MARKETABILITY

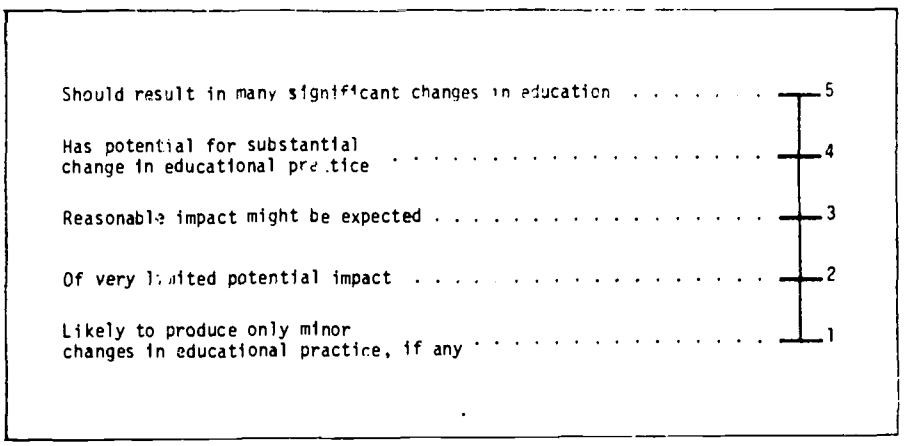


Comments: _____

_____

_____

K. POTENTIAL IMPACT



Comments and Explanations: _____

_____

_____

Date_____

Product Number_____

Evaluator_____

DEVELOPMENTAL PRODUCT RATING FORM

The following are abbreviated definitions of the criteria used to evaluate developmental products. More elaborate definitions are offered in the Evaluators' Manual.

A.	IMPORTANCE OF GENERAL PROBLEM:	. . . degree to which problem is crucial to education . . . magnitude of the problem
B.	RELEVANCE OF PRODUCT TO GENERAL PROBLEM:	. . . degree to which product clearly and directly relates to stated problem
C.	COMPREHENSIVENESS OF THE PRODUCT AS PROBLEM SOLUTION:	. . . degree to which product meets the whole problem
D.	CONTENT ACCURACY:	. . . informationally correct . . . a precise accounting and presentation
E.	CONTENT CLARITY:	. . . an easily understood exposition . . . full, unambiguous explanations and directions
F.	EFFECTIVENESS:	. . . degree to which product solves the problem . . . degree to which product meets its objectives
G.	REASONABLE COST TO ADOPT/ IMPLEMENT, GIVEN OUTCOME:	. . . degree to which product is worth buying, given what might or will come of its use
H.	REASONABLE COST TO USE/ OPERATE, GIVEN OUTCOME:	. . . degree to which product is worth continuing to use
I.	SCOPE OF POSSIBLE MARKET:	. . . possible number of users, buyers, clients
J.	AMENABILITY TO MARKETING:	. . . attractiveness of product . . . ease of acquisition and use
K.	POTENTIAL IMPACT:	. . . likelihood of effecting change in educational practices, given all factors

4-72/IID

INSTRUCTIONS

Your evaluation on each of the following criteria will be the result of a two step decision process. In the first step a fairly gross decision will be made. During the second step your initial decision will be further refined.

For example, if for criterion A you feel the problem addressed by the product is "among the most important in education today," you would select phrase 1. You would then consider just how important you really think it is. Is it of "critical" importance, or just "very" important? If the former, you would select "a," if the latter, you would choose "b."

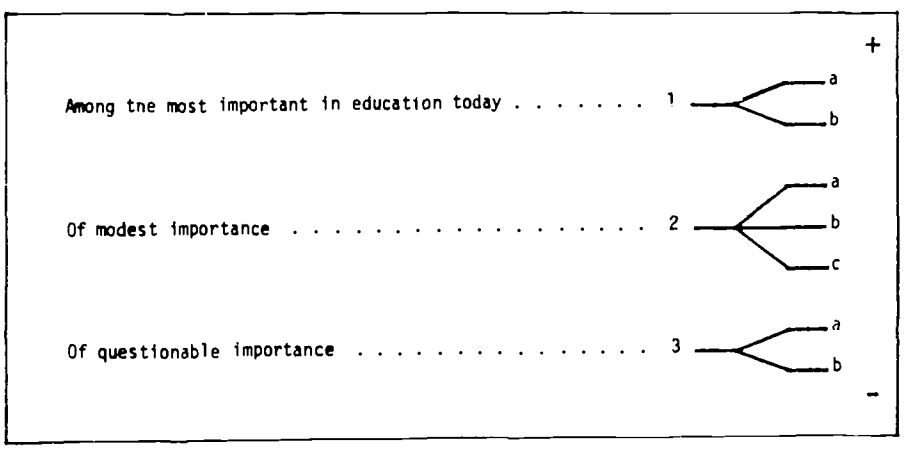
If you feel, however, the problem is only "of modest importance," you would then consider just how "moderate" you think the importance to be: above average, just average, or somewhat below average. You would then select "a," "b," or "c" accordingly.

If you feel the product is "of questionable importance," decide whether its importance is only questionable, or whether the product is of absolutely no importance at all, as far as you are concerned. If the former, you would select "a," if the latter, "b."

When you have made your judgment, circle the letter of your final decision.

Should you, for some reason, be unable to arrive at a rating on a particular criterion, note this and explain why in the Comments section. Also use the Comments sections for any additional remarks you may wish to make. Comments explaining very low ratings will be especially helpful. For the final criterion, Potential Impact, please explain why you feel the product will or will not have impact on the educational community.

A. PROBLEM IMPORTANCE

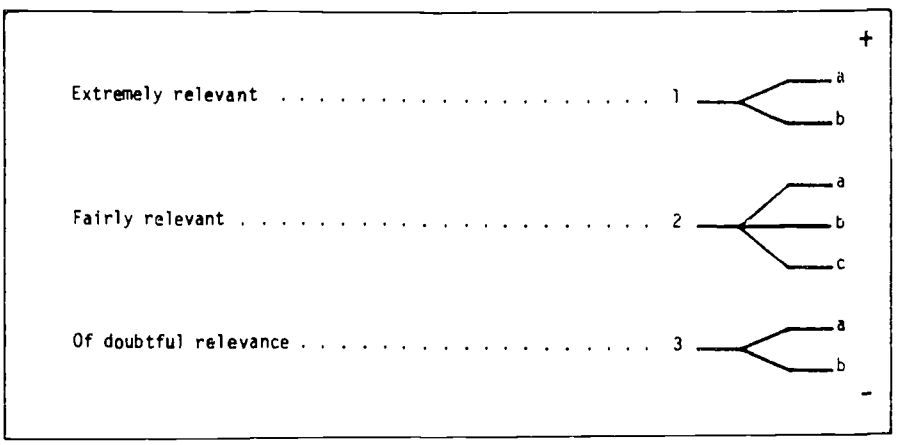


Comments: _____

_____

_____

B. RELEVANCE OF PRODUCT TO GENERAL PROBLEM

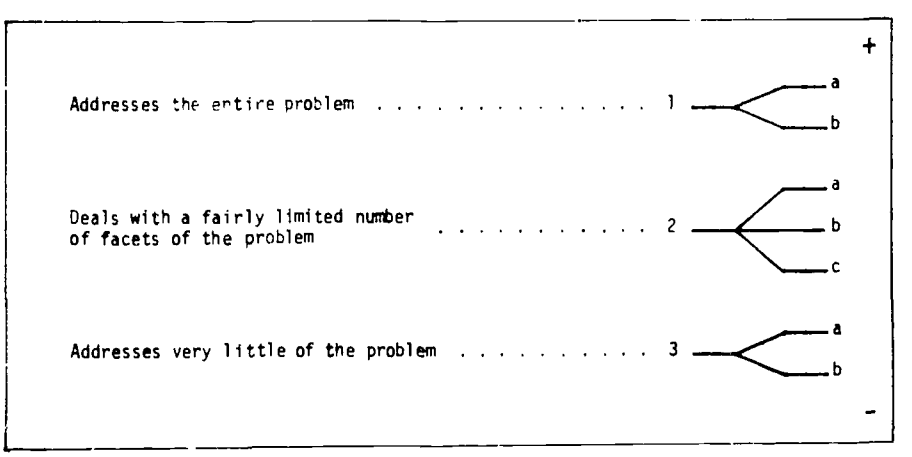


Comments: _____

_____

_____

C. COMPREHENSIVENESS OF THE PRODUCT AS PROBLEM SOLUTION

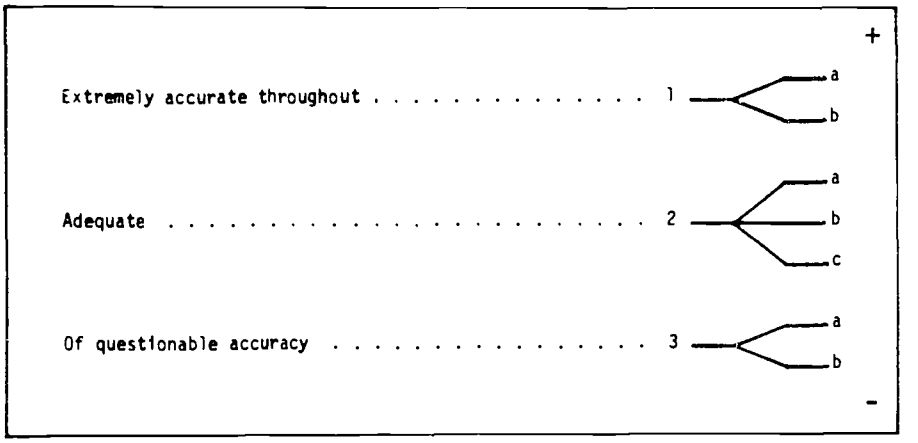


Comments: _____

_____

_____

D. CONTENT ACCURACY

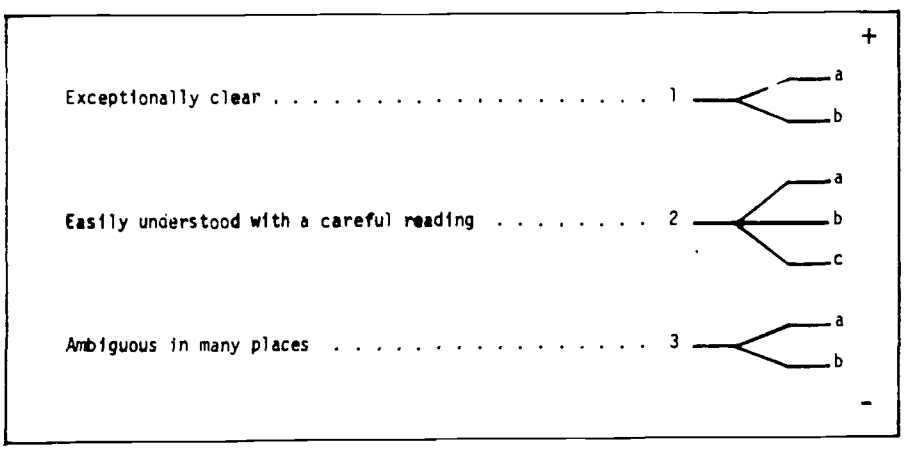


Comments: _____

_____

_____

E. CONTENT CLARITY



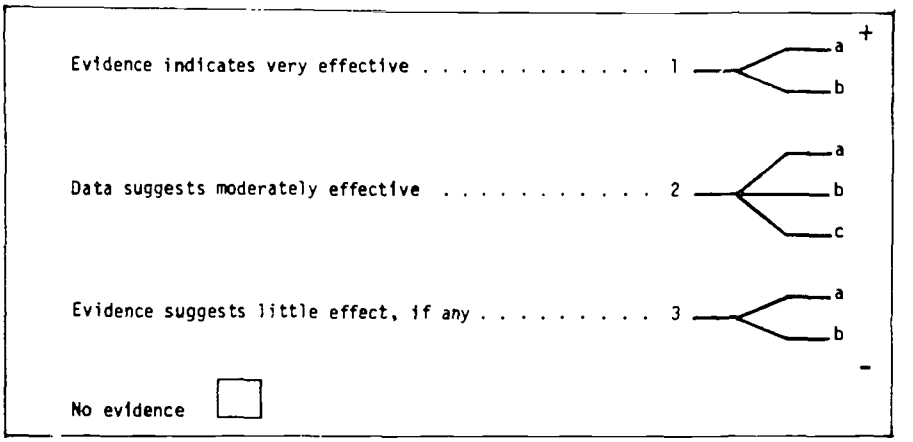
Comments: _____

_____

_____

F. EFFECTIVENESS

Note: if there is no evidence on which to judge the effectiveness of this product, indicate by checking the box labeled "No Evidence."

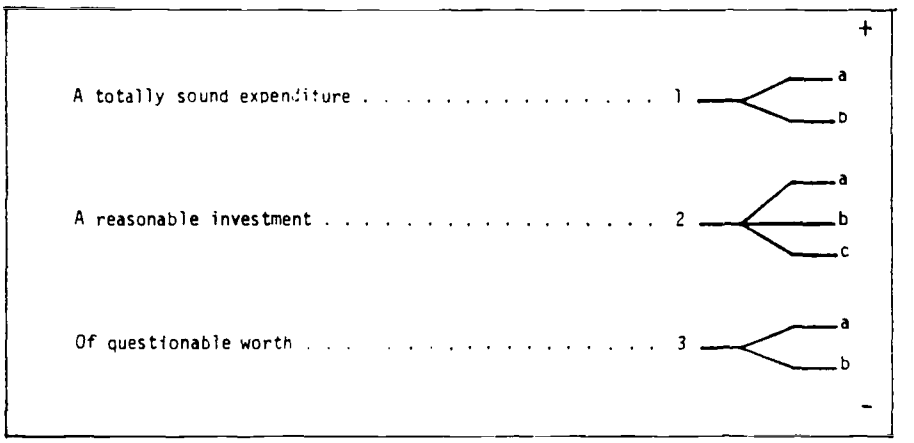


Comments: _____

_____

_____

G. REASONABLE COST TO ADOPT/IMPLEMENT, GIVEN OUTCOME

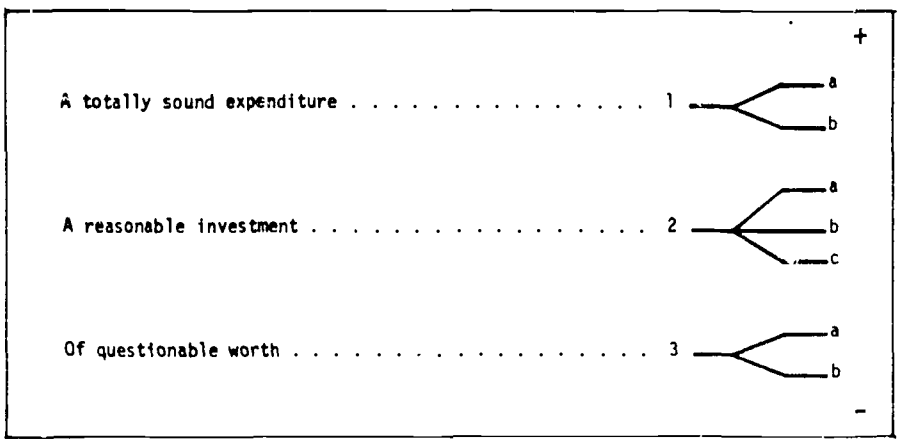


Comments: _____

_____

_____

H. REASONABLE COST TO USE/OPERATE, GIVEN OUTCOME

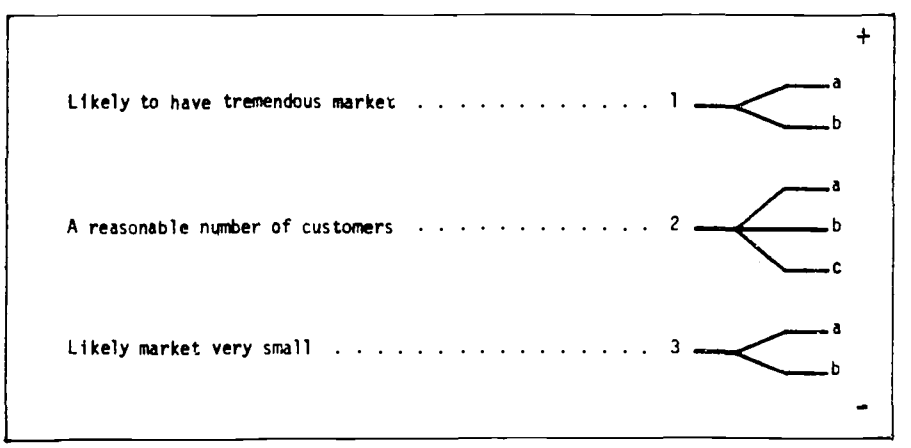


Comments: _____

_____

_____

I. POTENTIAL MARKET

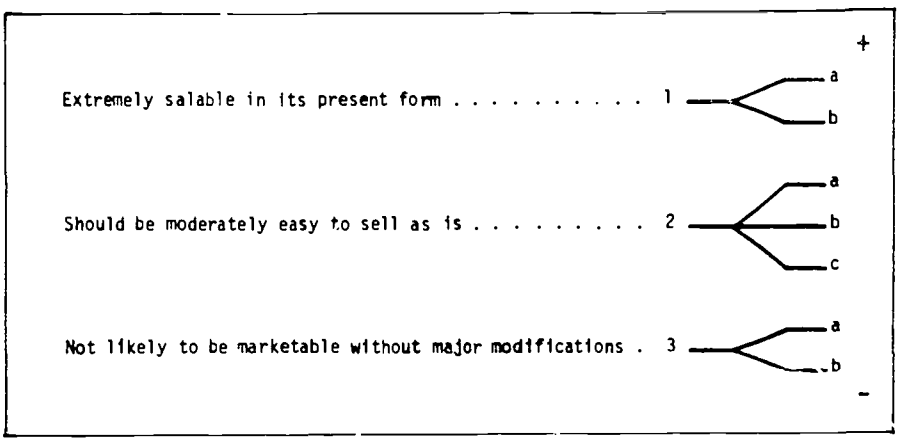


Comments: _____

_____

_____

J. POTENTIAL MARKETABILITY

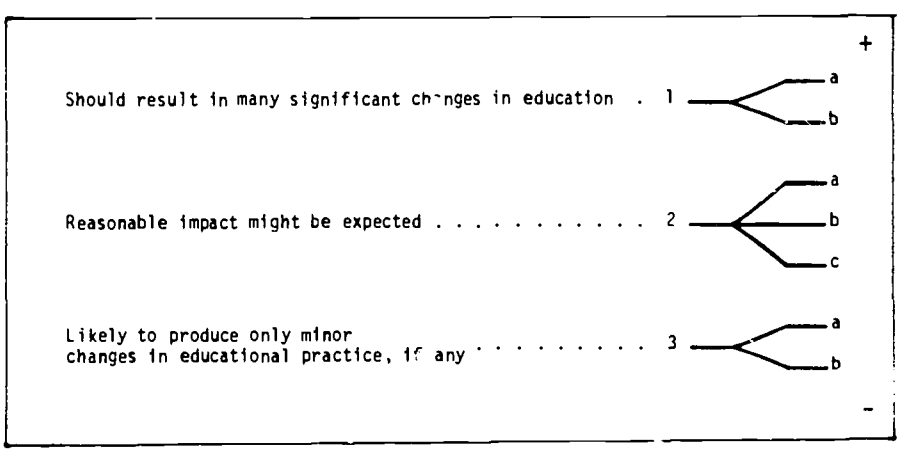


Comments: _____

_____

_____

K. POTENTIAL IMPACT



Comments and Explanations: _____

_____

_____

Date _____

Evaluator _____

Product Number _____

KNOWLEDGE PRODUCT RATING FORM

The following are abbreviated definitions of the criteria used to evaluate knowledge products. More elaborate definitions are offered in the Evaluators' Manual.

A.	IMPORTANCE OF GENERAL PROBLEM:	. . . degree to which problem is crucial to education . . . magnitude of the problem
B.	RELEVANCE OF PRODUCT TO GENERAL PROBLEM:	. . . degree to which product clearly and directly relates to stated problem
C.	COMPREHENSIVENESS OF THE PRODUCT AS PROBLEM SOLUTION:	. . . degree to which product meets the whole problem
D.	ORIGINALITY OF PRODUCT:	. . . extent to which product represents a unique contribution
E.	QUALITY OF LITERATURE DISCUSSION:	. . . exhibits an awareness of current "state of the art" . . . appropriate to problem area
F.	ADEQUACY OF RESEARCH DESIGN:	. . . appropriateness of statistical treatments . . . representativeness of sample
G.	APPROPRIATENESS OF INTERPRETATION:	. . . justified by the data
H.	REASONABLENESS OF CONCLUSIONS/ RECOMMENDATIONS:	. . . generally logical . . . substantiated by the findings
I.	CLARITY OF PRESENTATION:	. . . an easily understood exposition . . . full, unambiguous discussion
J.	POTENTIAL IMPACT:	. . . likelihood of effecting change in educational practices, given all factors

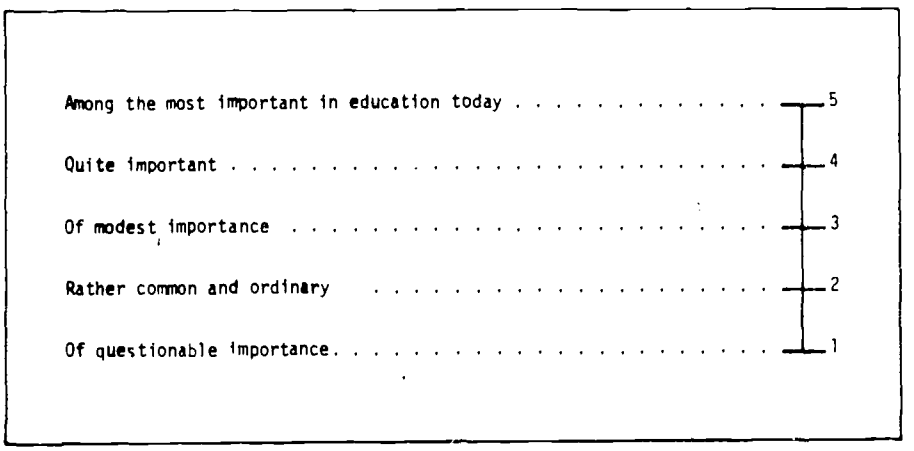
4-72/IK

INSTRUCTIONS

For each scale, select that phrase which best represents your judgment of the product. Then circle the number of that phrase. Do not mark intermediate points.

Should you, for some reason, be unable to arrive at a rating on a particular criterion, note this and explain why in the Comments section. Also use the Comments sections for any additional remarks you may wish to make. Comments explaining very low ratings will be especially helpful. For the final criterion, Potential Impact, please explain why you feel the product will or will not have impact on the educational community.

A. PROBLEM IMPORTANCE

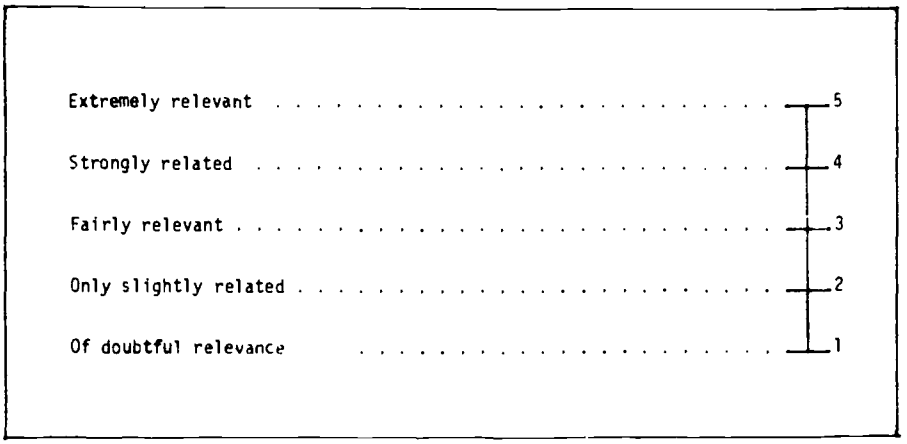


Comments: _____

_____

_____

B. RELEVANCE OF PRODUCT TO PROBLEM

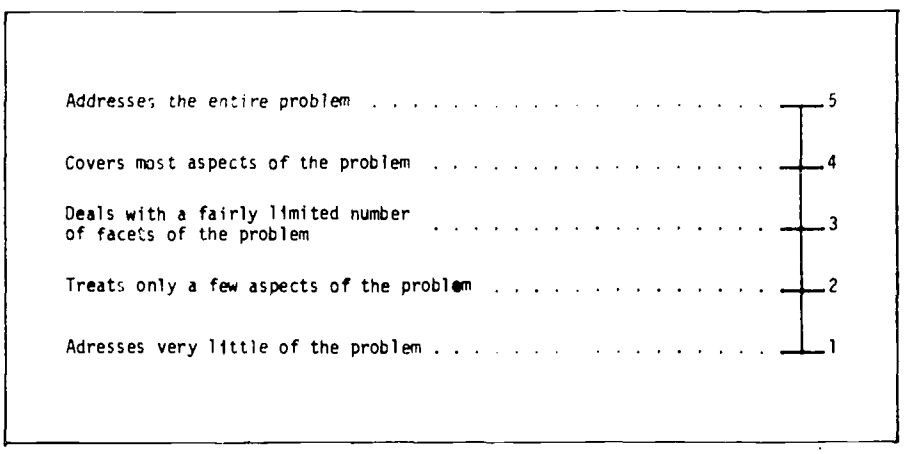


Comments: _____

_____

_____

C. COMPREHENSIVENESS OF PRODUCT AS PROBLEM SOLUTION

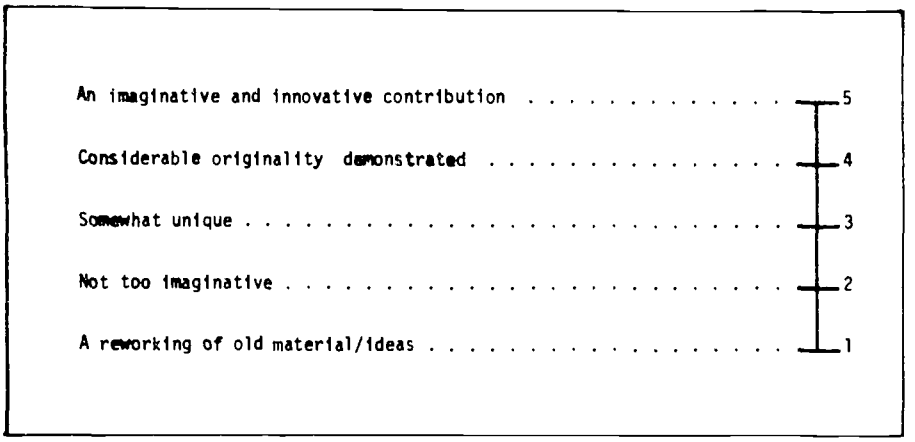


Comments: _____

_____

_____

D. ORIGINALITY



Comments: _____

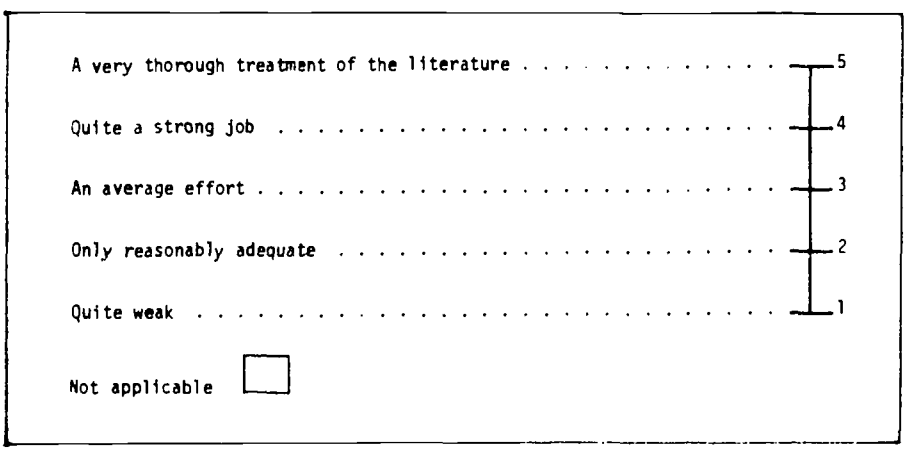
_____

_____

* * * * *

Note: the following four criteria may not be appropriate for all knowledge products. For example, all knowledge products do not necessarily contain a review of the literature. If any of the next four criteria is inappropriate for the product being evaluated, please indicate by checking the box labeled "Not Applicable" for that criterion.

E. QUALITY OF LITERATURE DISCUSSION

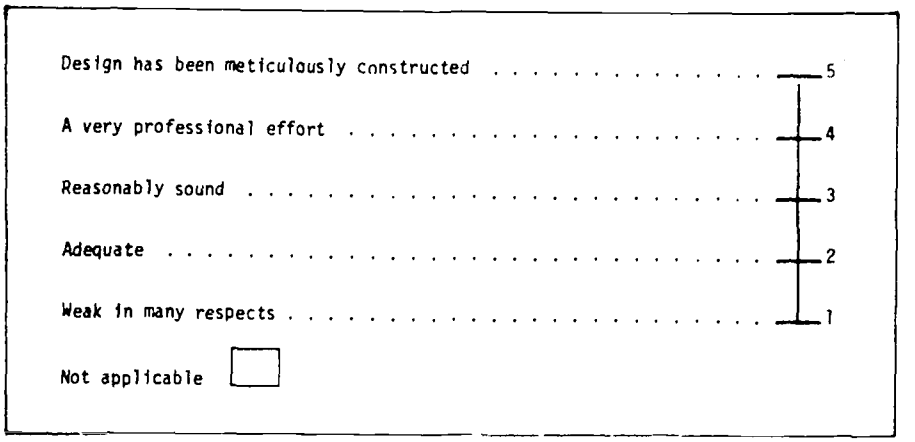


Comments: _____

_____

_____

F. ADEQUACY OF RESEARCH DESIGN

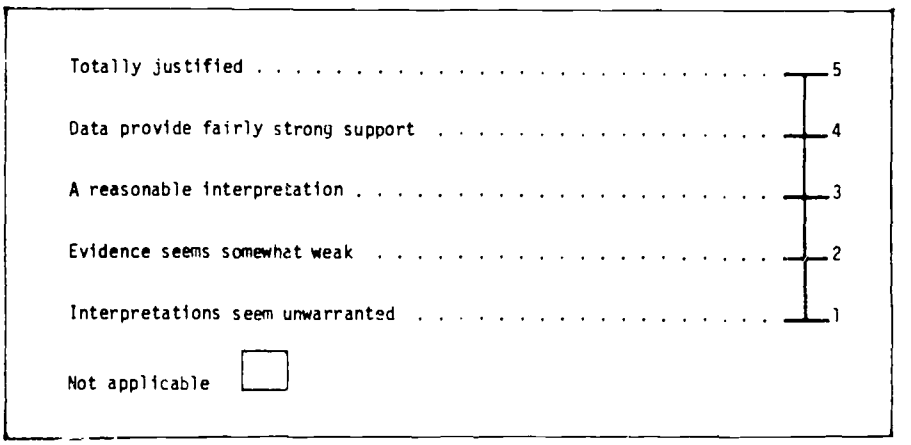


Comments: _____

_____

_____

G. APPROPRIATENESS OF INTERPRETATIONS

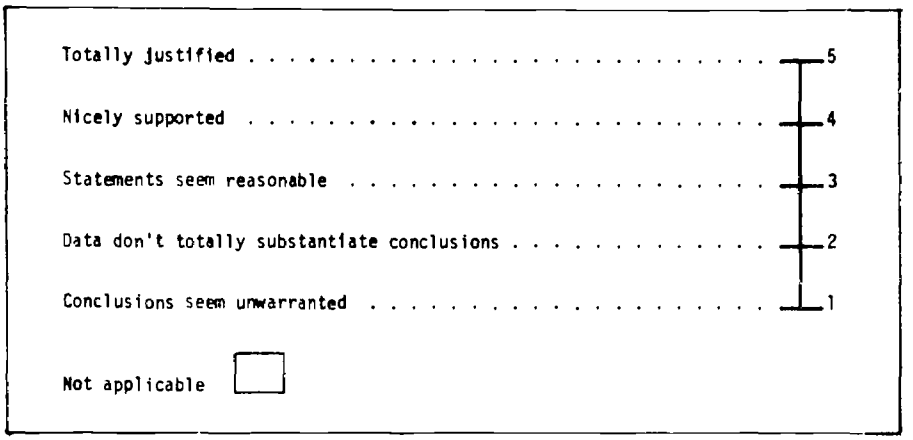


Comments: _____

_____

_____

H. REASONABLENESS OF CONCLUSIONS/RECOMMENDATIONS



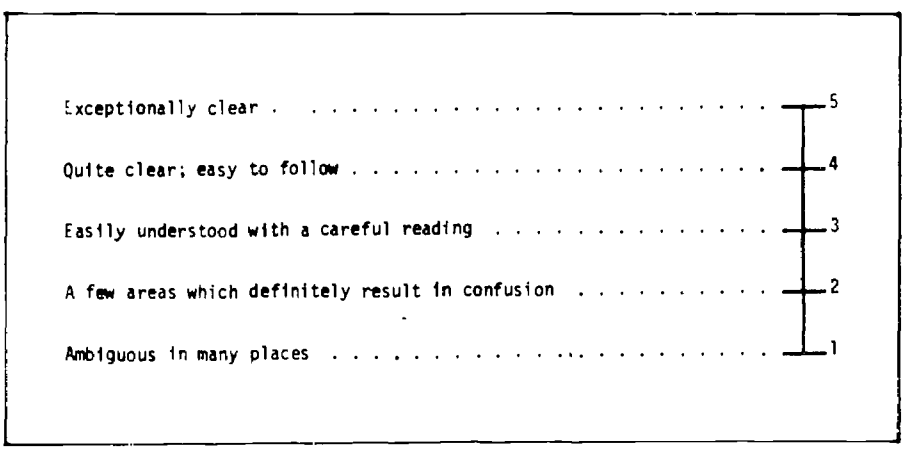
Comments: _____

_____

_____

* * * * *

I. CLARITY OF PRESENTATION

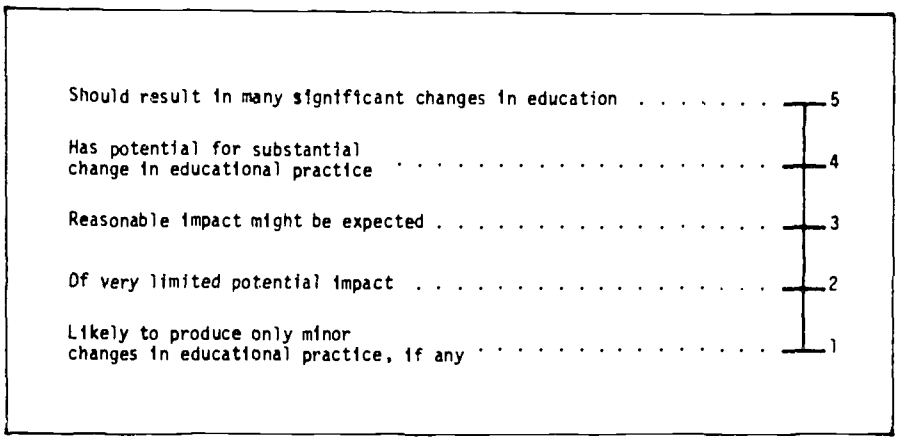


Comments: _____

_____

_____

J. POTENTIAL IMPACT



Comments and Explanations: _____

_____

_____

Date_____

Evaluator_____

Product Number_____

KNOWLEDGE PRODUCT RATING FORM

The following are abbreviated definitions of the criteria used to evaluate knowledge products. More elaborate definitions are offered in the Evaluators' Manual.

A.	IMPORTANCE OF GENERAL PROBLEM:	. . . degree to which problem is crucial to education . . . magnitude of the problem
B.	RELEVANCE OF PRODUCT TO GENERAL PROBLEM:	. . . degree to which product clearly and directly relates to stated problem
C.	COMPREHENSIVENESS OF THE PRODUCT AS PROBLEM SOLUTION:	. . . degree to which product meets the whole problem
D.	ORIGINALITY OF PRODUCT:	. . . extent to which product represents a unique contribution
E.	QUALITY OF LITERATURE DISCUSSION:	. . . exhibits an awareness of current "state of the art" . . . appropriate to problem area
F.	ADEQUACY OF RESEARCH DESIGN:	. . . appropriateness of statistical treatments . . . representativeness of sample
G.	APPROPRIATENESS OF INTERPRETATION:	. . . justified by the data
H.	REASONABLENESS OF CONCLUSIONS/ RECOMMENDATIONS:	. . . generally logical . . . substantiated by the findings
I.	CLARITY OF PRESENTATION:	. . . an easily understood exposition . . . full, unambiguous discussion
J.	POTENTIAL IMPACT:	. . . likelihood of effecting change in educational practices, given all factors

4-72/IIK

INSTRUCTIONS

Your evaluation on each of the following criteria will be the result of a two step decision process. In the first step a fairly gross decision will be made. During the second step your initial decision will be further refined.

For example, if for criterion A you feel the problem addressed by the product is "among the most important in education today," you would select phrase 1. You would then consider just how important you really think it is. Is it of "critical" importance, or just "very" important? If the former, you would select "a," if the latter, you would choose "b."

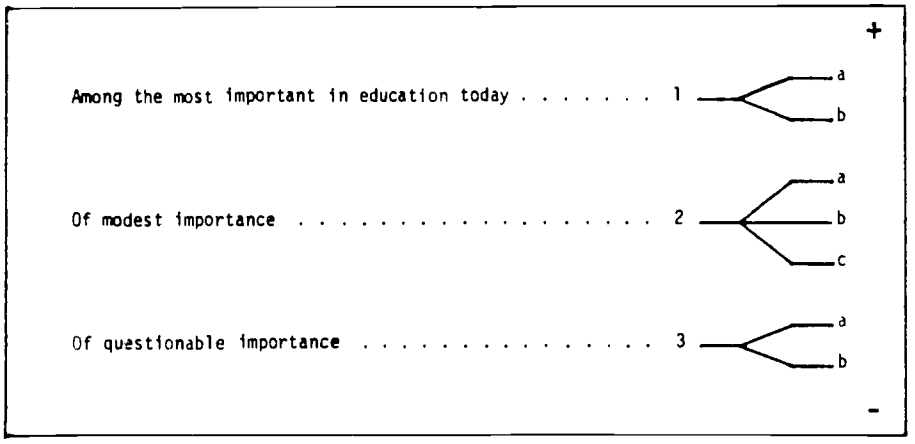
If you feel, however, the problem is only "of modest importance," you would then consider just how "moderate" you think the importance to be: above average, just average, or somewhat below average. You would then select "a," "b," or "c" accordingly.

If you feel the product is "of questionable importance," decide whether its importance is only questionable, or whether the product is of absolutely no importance at all, as far as you are concerned. If the former, you would select "a," if the latter, "b."

When you have made your judgment, circle the letter of your final decision.

Should you, for some reason, be unable to arrive at a rating on a particular criterion, note this and explain why in the Comments section. Also use the Comments sections for any additional remarks you may wish to make. Comments explaining very low ratings will be especially helpful. For the final criterion, Potential Impact, please explain why you feel the product will or will not have impact on the educational community.

A. PROBLEM IMPORTANCE

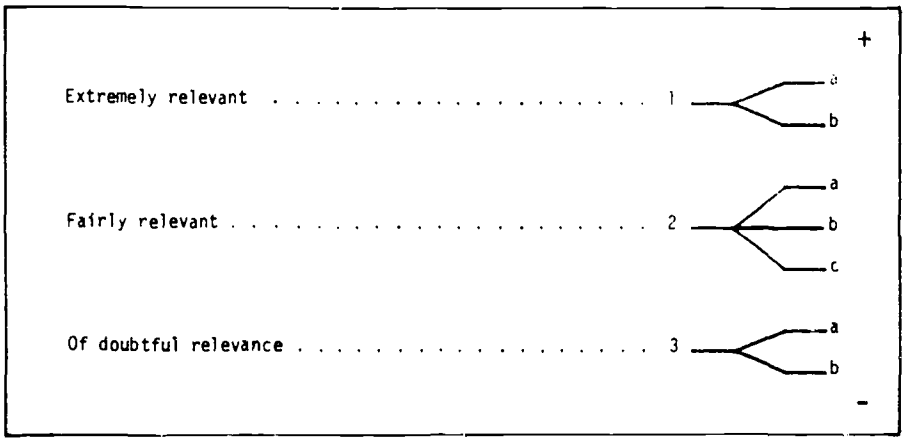


Comments: _____

_____

_____

B. RELEVANCE OF PRODUCT TO PROBLEM

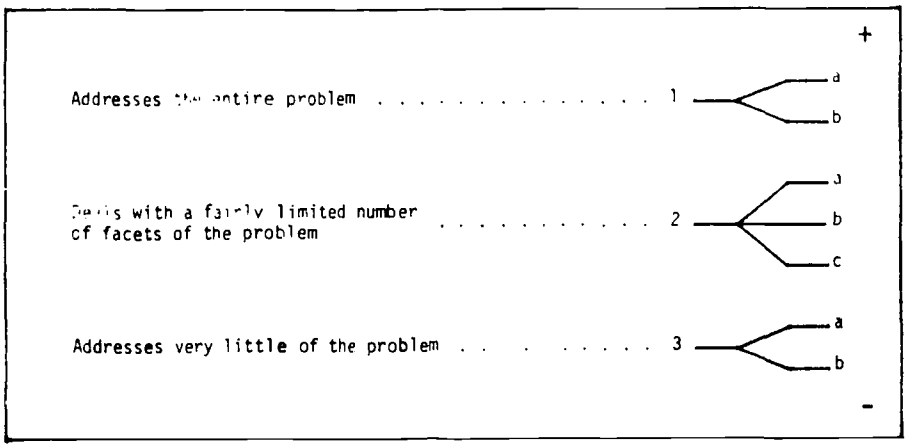


Comments: _____

_____

_____

C. COMPREHENSIVENESS OF PRODUCT AS PROBLEM SOLUTION

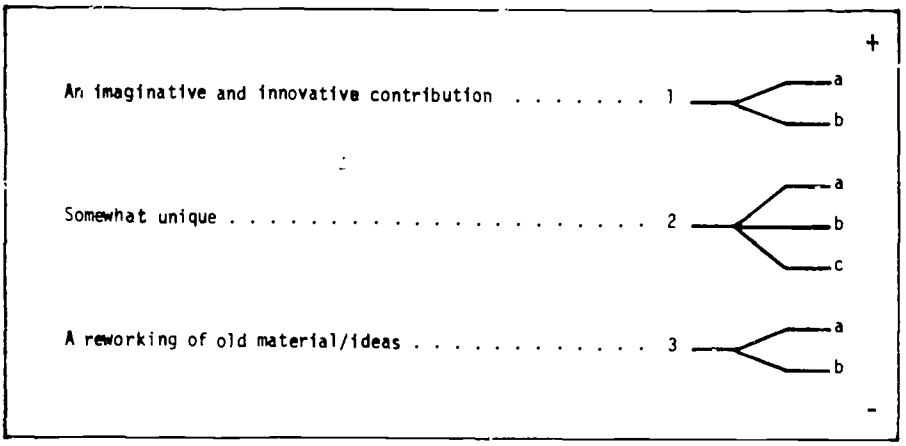


Comments: _____

_____

_____

D. ORIGINALITY



Comments: _____

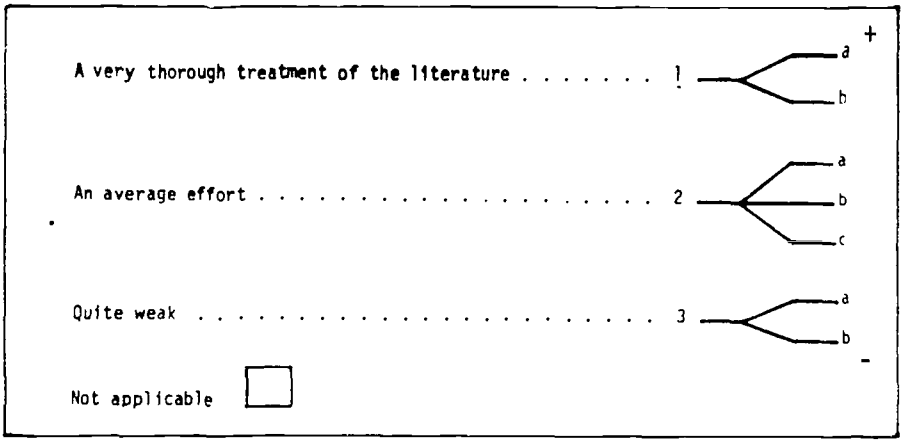
_____

_____

* * * * *

Note: the following four criteria may not be appropriate for all knowledge products. For example, all knowledge products do not necessarily contain a review of the literature. If any of the next four criteria is inappropriate for the product being evaluated, please indicate by checking the box labeled "Not Applicable" for that criterion.

E. QUALITY OF LITERATURE DISCUSSION

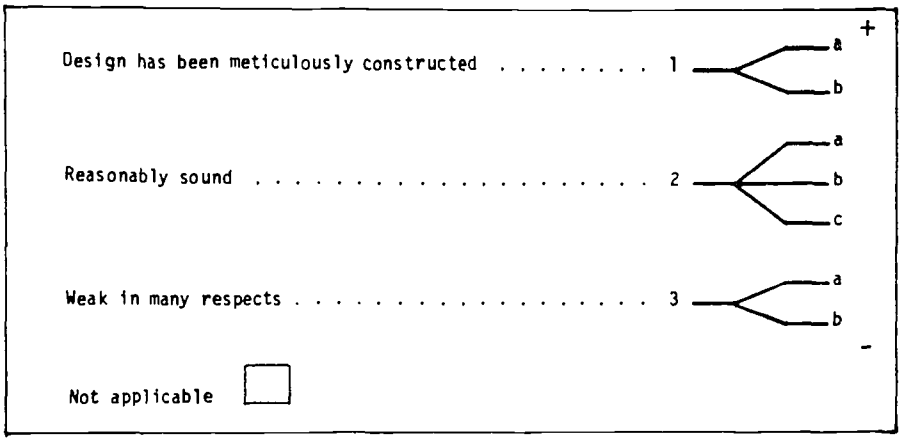


Comments: _____

_____

_____

F. ADEQUACY OF RESEARCH DESIGN

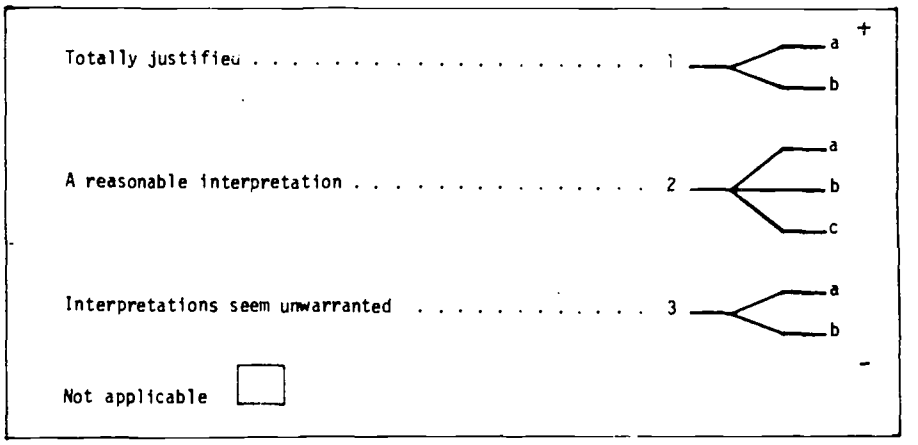


Comments: _____

_____

_____

G. APPROPRIATENESS OF INTERPRETATIONS

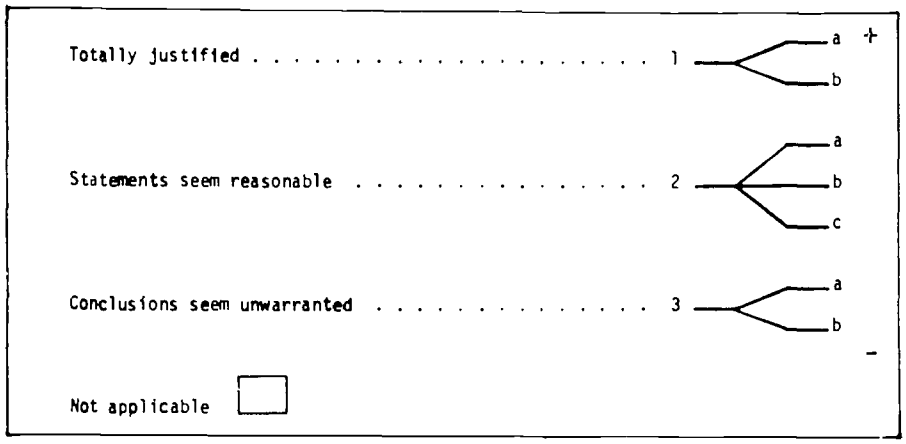


Comments: _____

_____

_____

H. REASONABLENESS OF CONCLUSIONS/RECOMMENDATIONS



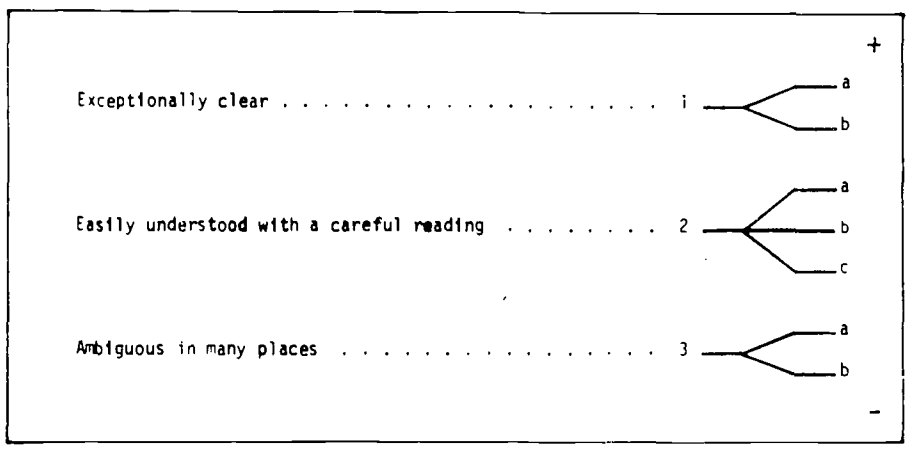
Comments: _____

_____

_____

* * * * *

I. CLARITY OF PRESENTATION

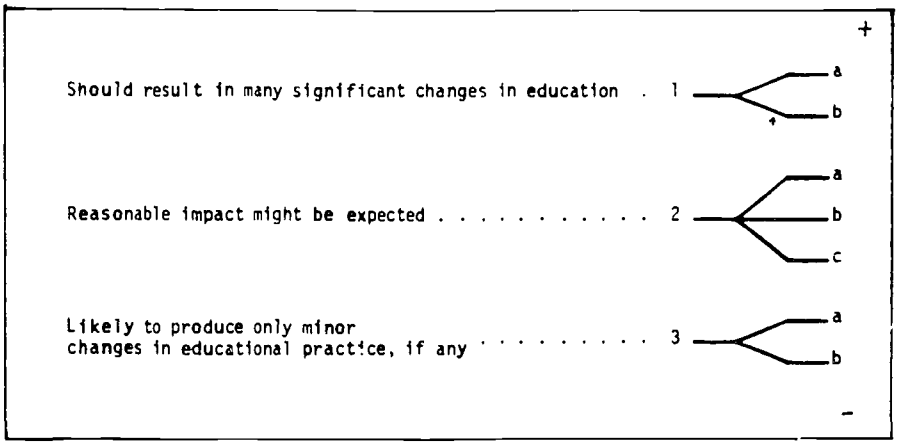


Comments: _____

_____

_____

J. POTENTIAL IMPACT



Comments and Explanations: _____

_____

_____

Appendix C

EVALUATION DATA SUMMARY SHEETS

DEVELOPMENTAL PRODUCTS RATING SUMMARY SHEET

Date _____

Product Number _____

Product/Panel Area _____

_____

Evaluator	A. Problem Importance	B. Relevance of Product to General Problem	C. Comprehensiveness of Product as Problem Solution	D. Content Accuracy	E. Content Clarity	F. Effectiveness	G. Reasonable Cost to Adopt/Implement, Given Outcome	H. Reasonable Cost to Use/Operate, Given Outcome	I. Potential Market	J. Potential Marketability	K. Potential Impact	Special Comments
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												

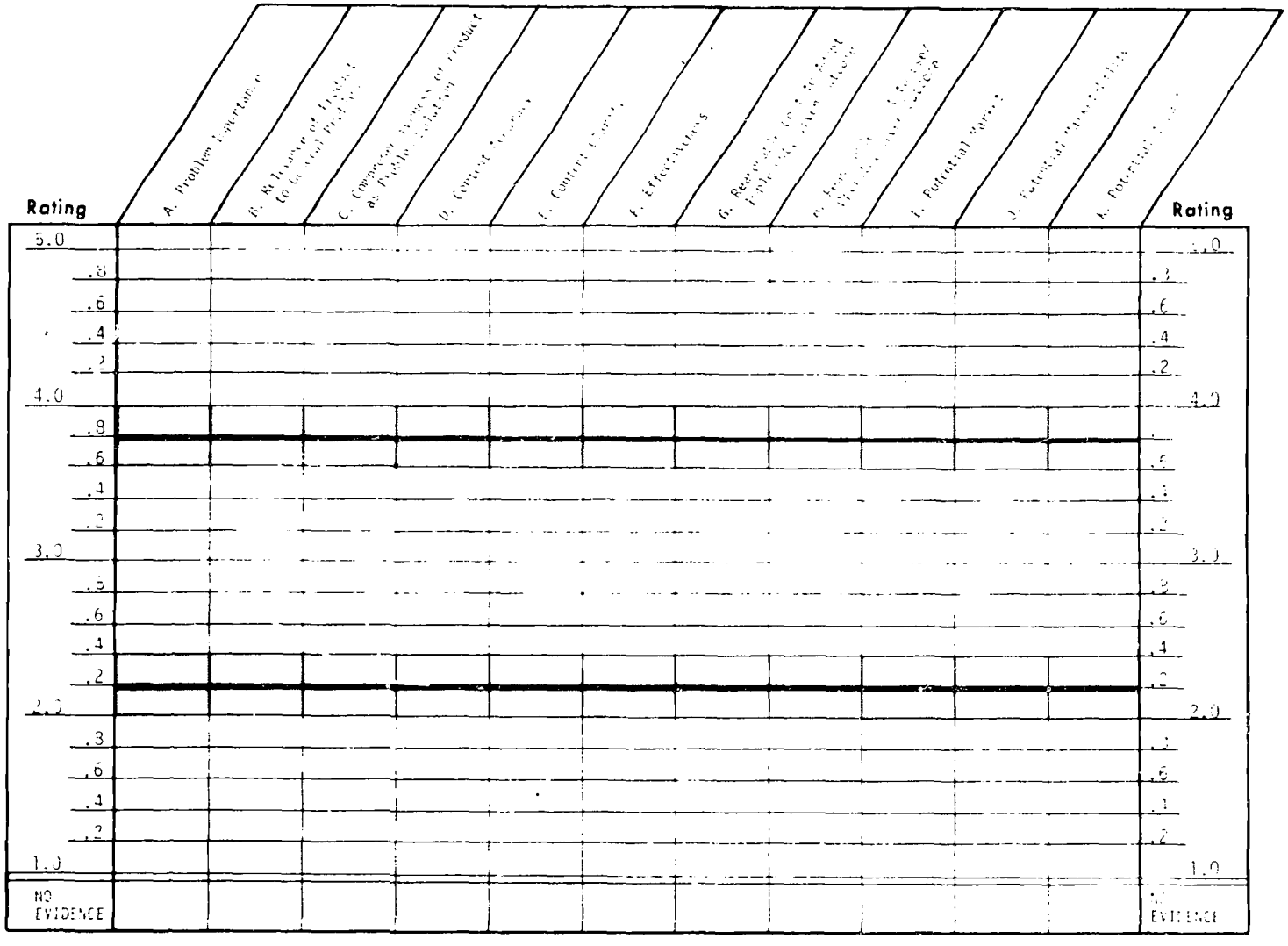
$\bar{X}$												
Range												

DEVELOPMENTAL PRODUCTS EVALUATION SUMMARY SHEET

DATE _____ PRODUCT AREA _____

NUMBER OF EVALUATORS _____ _____

Criteria



PRODUCT CODE

A ___

B ___

C ___

D ___

E ___

F ___

G ___

H ___

I ___

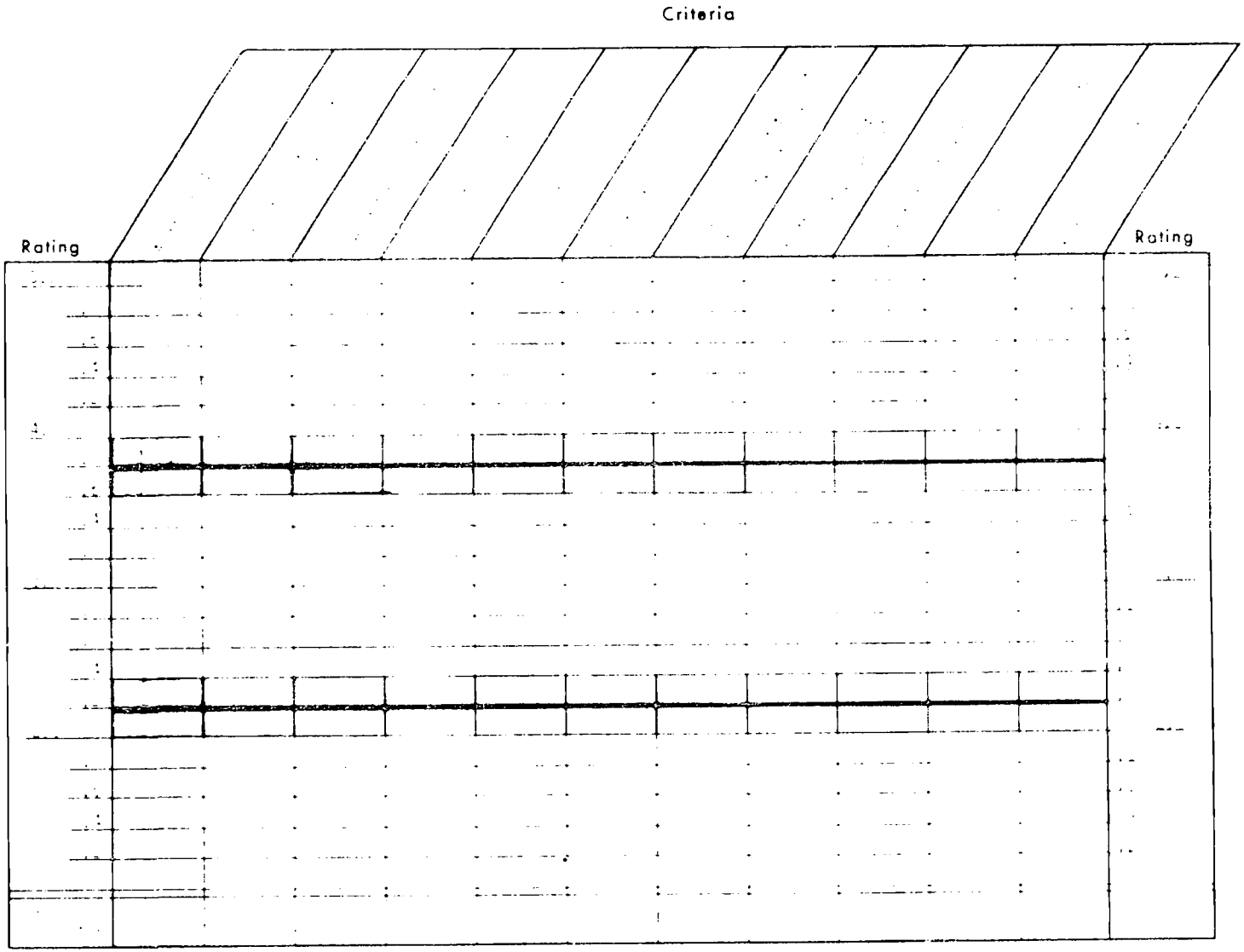
J ___

DEVELOPMENTAL PRODUCTS MULTIPLE PROFILES SHEET

DATE _____

PRODUCT AREA _____

NUMBER OF EVALUATORS _____



PRODUCT CODE

A

B

C

D

E

F

G

H

I

J

KNOWLEDGE PRODUCTS RATING SUMMARY SHEET

Date _____

Product Number _____

Product/Panel Area _____

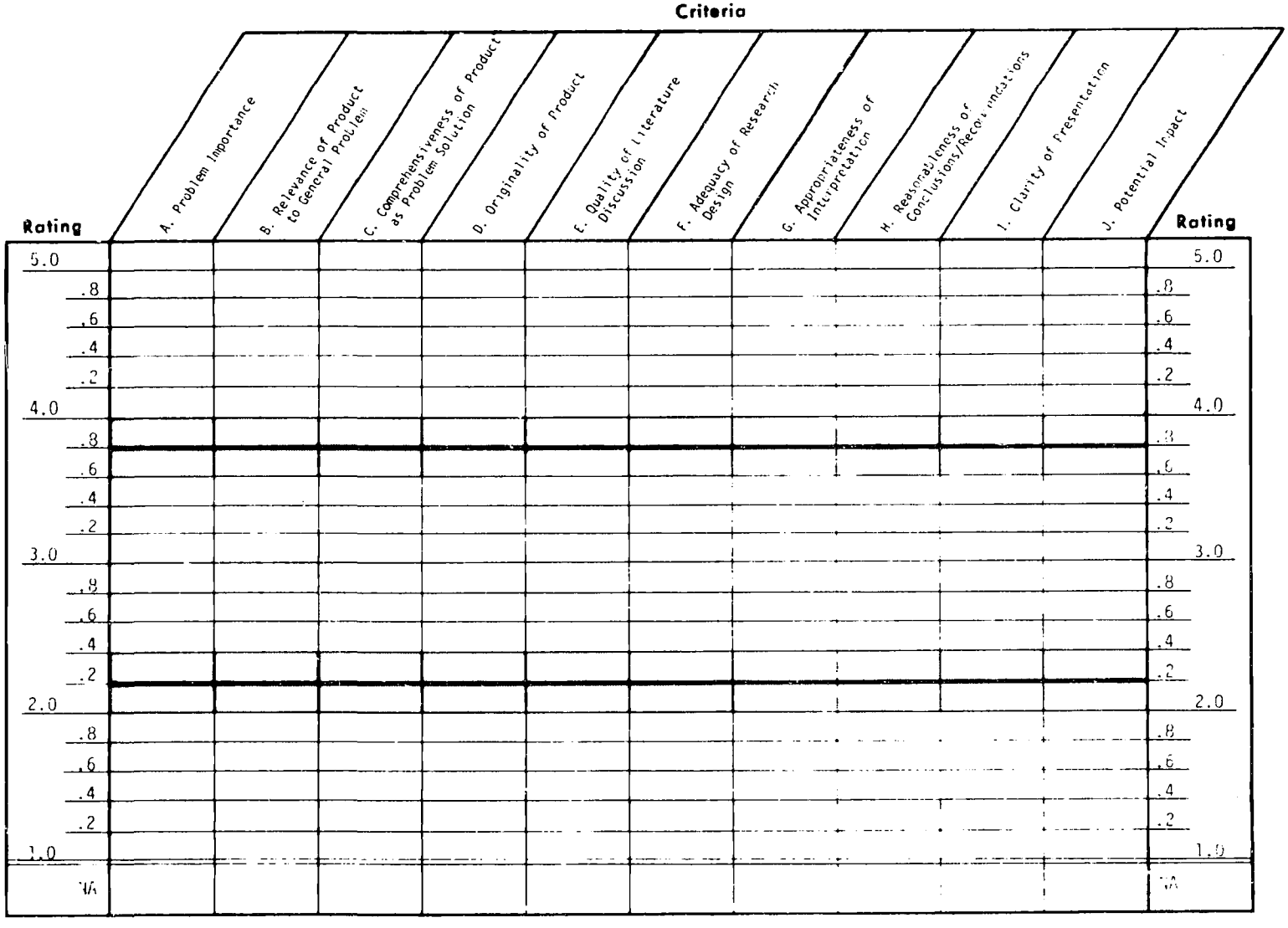
Evaluator	Criteria: A. Problem Importance	B. Relevance of Product to General Problem	C. Comprehensiveness of Product as Problem Solution	D. Originality of Product	E. Quality of Literature Discussion	F. Adequacy of Research Design	G. Appropriateness of Interpretation	H. Reasonableness of Conclusions/Recommendations	I. Clarity of Presentation	J. Potential Impact	Special Comments
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
$\bar{X}$											
Range											

KNOWLEDGE PRODUCTS EVALUATION SUMMARY SHEET

DATE _____

PRODUCT AREA _____

NUMBER OF EVALUATORS _____



PRODUCT CODE

A _____

B _____

C _____

D _____

E _____

F _____

G _____

H _____

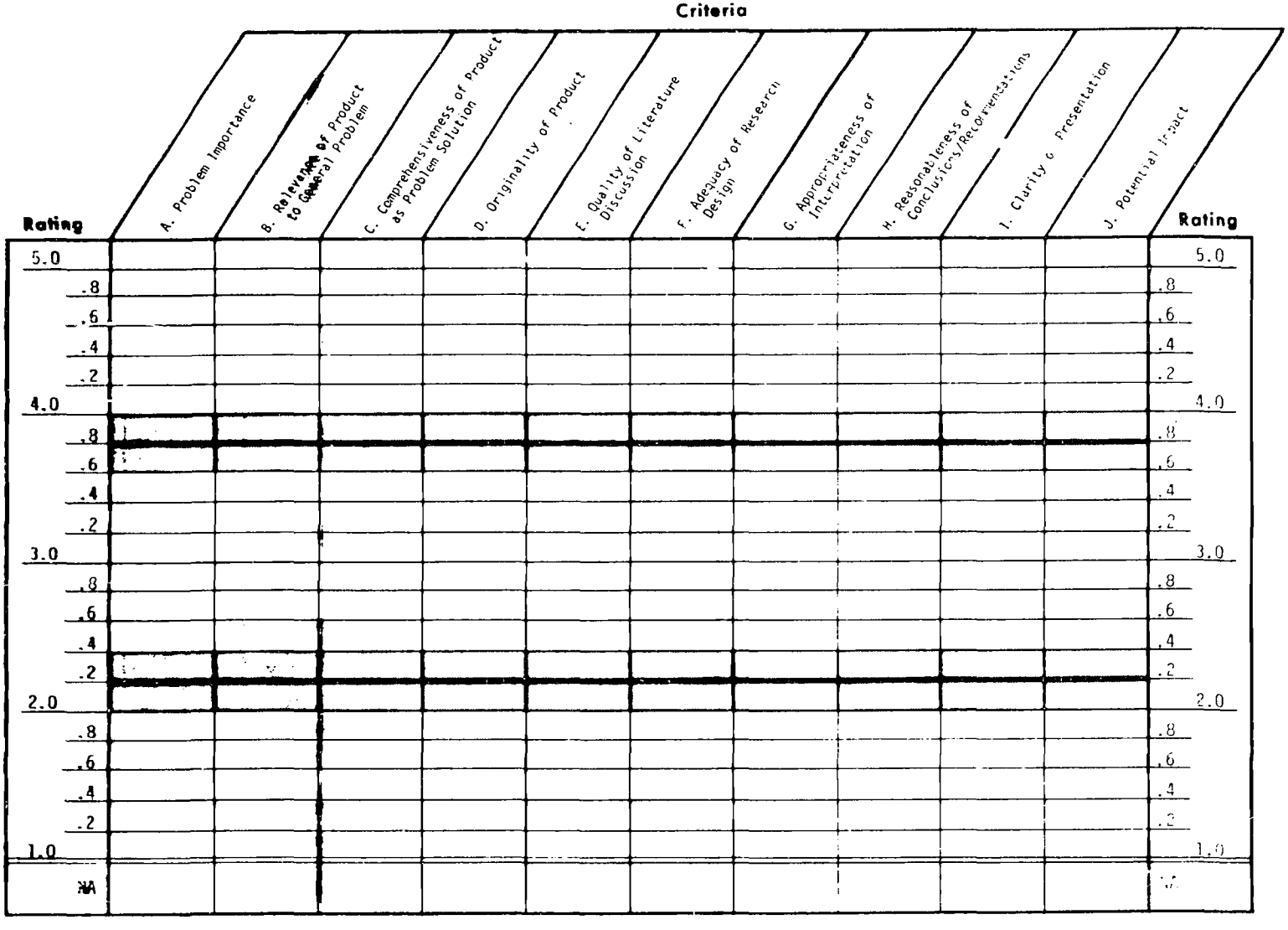
I _____

J _____

KNOWLEDGE PRODUCTS MULTIPLE PROFILES SHEET

DATE _____ **PRODUCT AREA** _____

NUMBER OF EVALUATORS _____



PRODUCT CODE

A _____

B _____

C _____

D _____

E _____

F _____

G _____

H _____

I _____

J _____

Appendix D

SAMPLE DATA SUMMARY SHEETS
RESULTING FROM THE PILOT TEST

DEVELOPMENTAL PRODUCTS RATING SUMMARY SHEET

Date 3 May 1972

Product Number A

Product/Panel Area The Learner and the Learning Process

Evaluator	A. Problem Importance	B. Relevance of Product to General Problem	C. Comprehensiveness of Product as Problem Solution	D. Content Accuracy	E. Content Clarity	F. Effectiveness	G. Reasonable Cost to Adopt/ Implement, Given Outcome	H. Reasonable Cost to Use/ Operate, Given Outcome	I. Potential Market	J. Potential Marketability	K. Potential Impact	Special Comments
	Criteria											
1	4	4	4	4	4	3	3	3	3	3	3	
2	4	5	4	4	4	3	3	3	3	3	3	
3	5	4	4	5	5	4	3	2	2	3	3	
4	4	5	4	4	4	3	4	5	2	3	4	Evaluator felt costs as stated were underestimated
5	5	5	5	5	5	3	-	3	3	-	4	Evaluator felt not provided sufficient information to judge start-up costs
6												
7												
8												
9												
10												
$\bar{X}$	4.4	4.6	4.2	4.4	4.4	3.2	3.3	3.2	2.6	3.0	3.4	
Range	4-5	4-5	4-5	4-5	4-5	3-4	3-4	2-5	2-3	3-3	3-4	

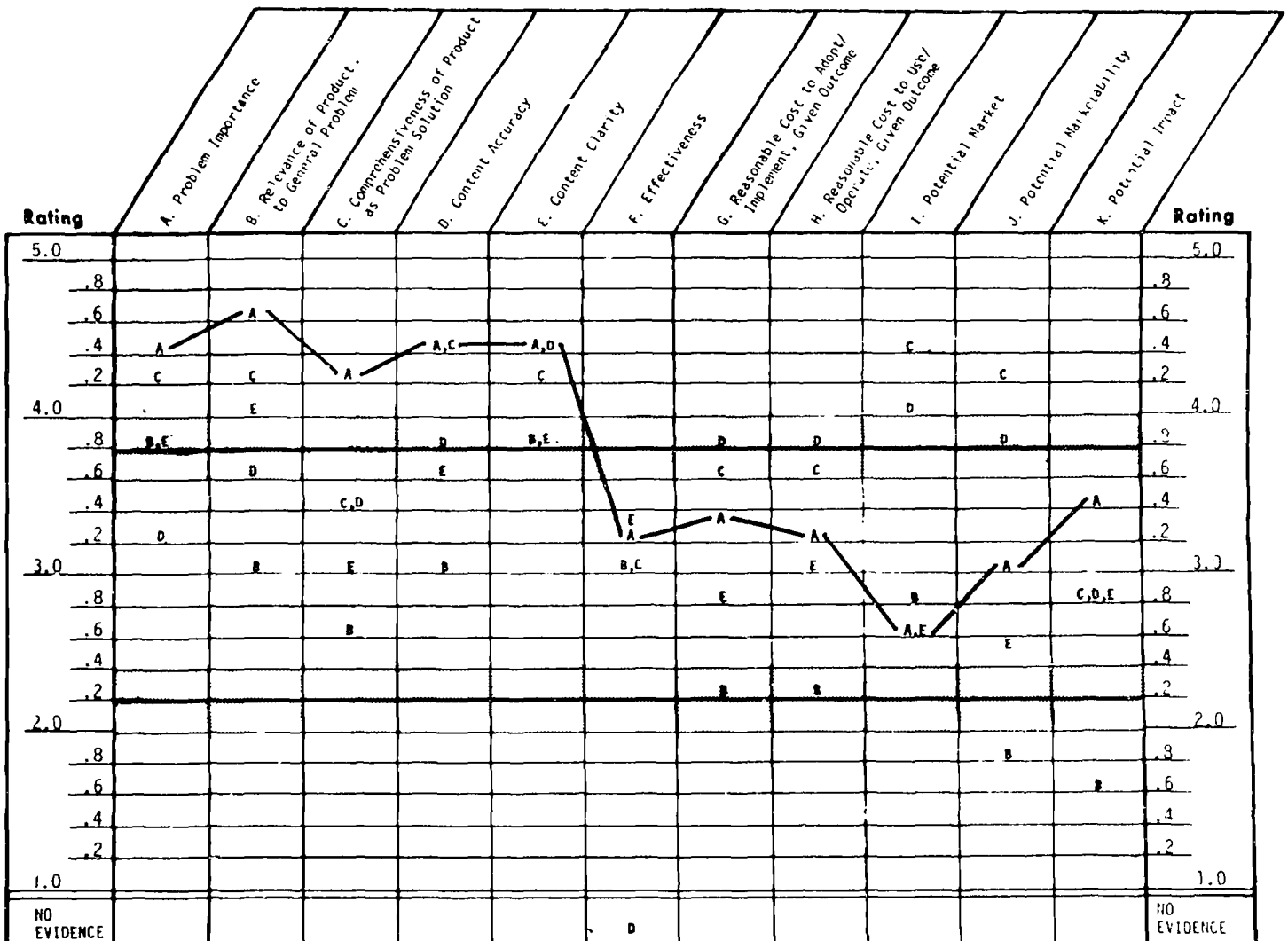
DEVELOPMENTAL PRODUCTS EVALUATION SUMMARY SHEET

DATE 3 May 1972

PRODUCT AREA The Learner and the Learning Process

NUMBER OF EVALUATORS 5

Criteria



The ratings of your product (identified above) in relation to the ratings of four other Developmental Product evaluated by the same panel

PRODUCT CODE

A ✓

B --

C --

D --

E --

F

G

H

I

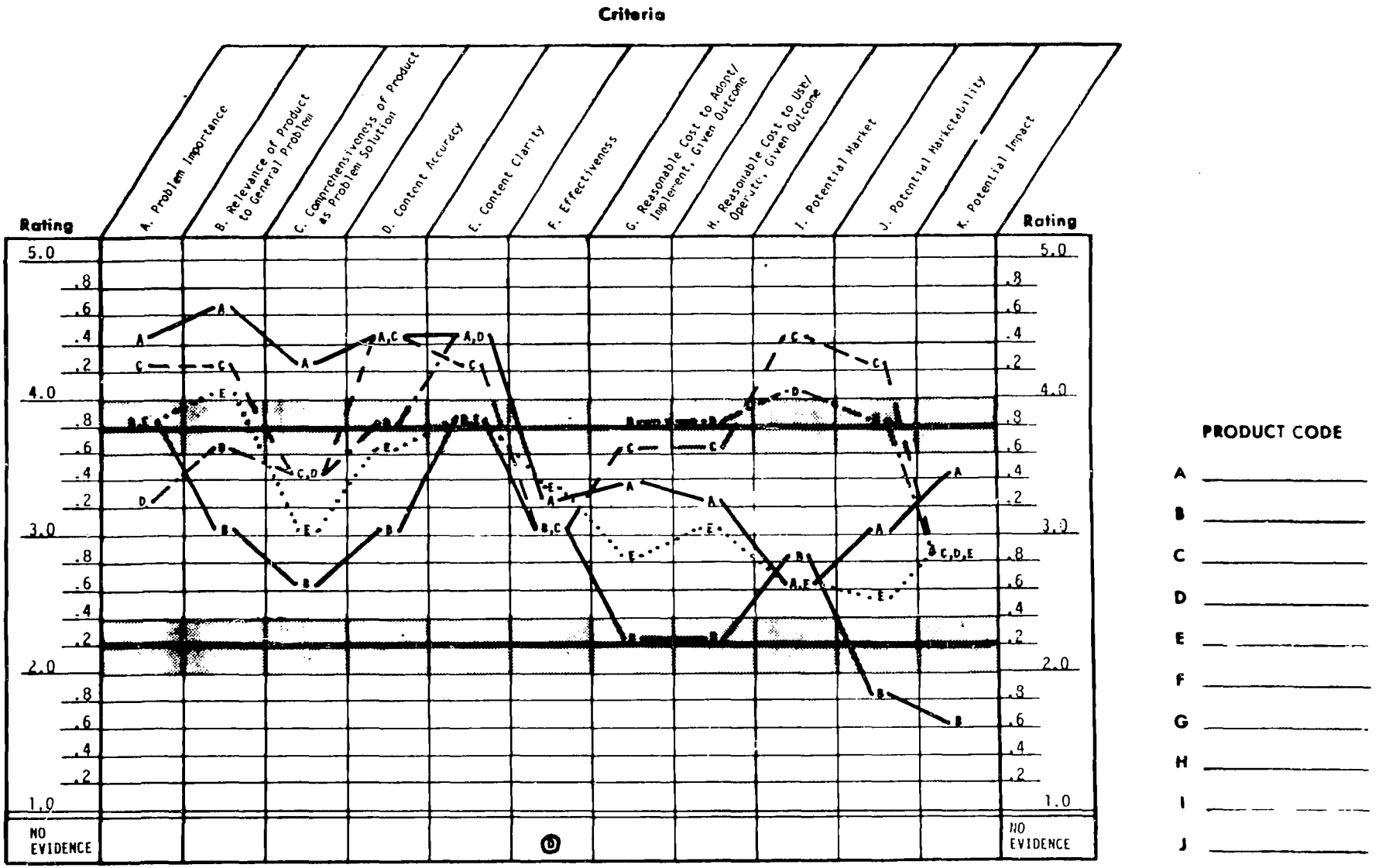
J

DEVELOPMENTAL PRODUCTS MULTIPLE PROFILES SHEET

DATE 3 May 1972

PRODUCT AREA The Learner and the Learning Process

NUMBER OF EVALUATORS 5



PRODUCT CODE

A _____

B _____

C _____

D _____

E _____

F _____

G _____

H _____

I _____

J _____

KNOWLEDGE PRODUCTS RATING SUMMARY SHEET

Date 3 May 1972

Product Number A

Product/Panel Area The Learner and the Learning Process

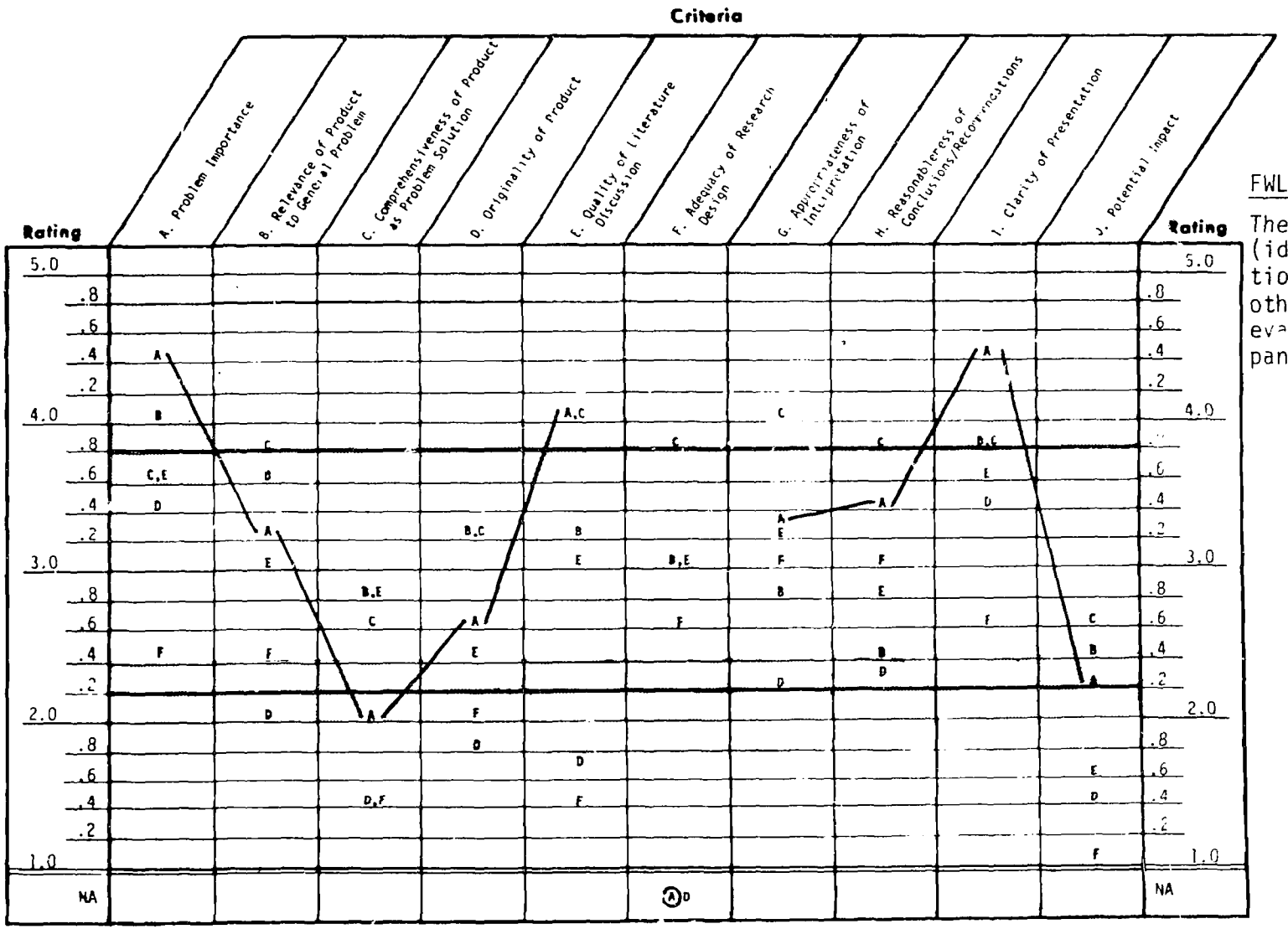
Evaluator	A. Problem Importance	B. Relevance of Product to General Problem	C. Comprehensiveness of Product as Problem Solution	D. Originality of Product	E. Quality of Literature Discussion	F. Adequacy of Research Design	G. Appropriateness of Interpretation	H. Reasonableness of Conclusions/Recommendations	I. Clarity of Presentation	J. Potential Impact	Special Comments
1	4	4	2	3	4	NA	3	3	4	3	
2	5	3	2	3	4	NA	3	3	4	2	
3	4	4	3	3	4	NA	4	3	5	3	Evaluator felt information would modify teacher's outlook;
4	5	3	2	2	4	NA	3	3	5	2	Evaluator felt more was needed than simply providing information;
5	4	2	1	2	4	NA	NA	5	4	1	Evaluator felt providing only advice/information would have little effect;
6											
7											
8											
9											
10											
$\bar{X}$	4.4	3.2	2.0	2.6	4.0	-	3.3	3.4	4.4	2.2	
Range	4-5	2-4	1-3	2-3	4	-	3-4	3-5	4-5	1-3	

KNOWLEDGE PRODUCTS EVALUATION SUMMARY SHEET

DATE 3 May 1972

PRODUCT AREA The Learner and the Learning Process

NUMBER OF EVALUATORS 5



FWLERD

The ratings of your product (identified above) in relation to the ratings of five other Knowledge Products evaluated by the same panel.

PRODUCT CODE

A ✓

B --

C --

D --

E --

F --

G

H

I

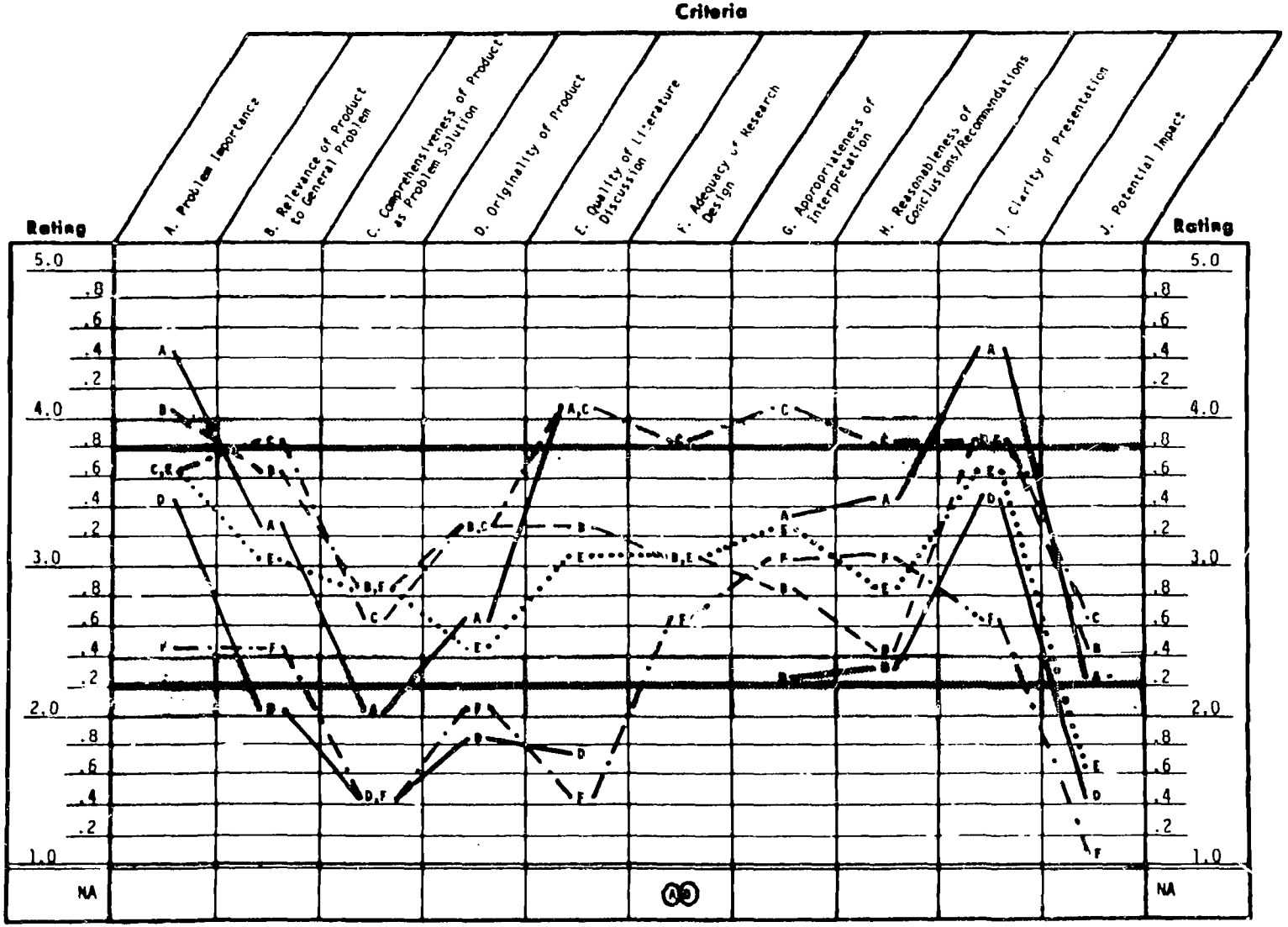
J

KNOWLEDGE PRODUCTS MULTIPLE PROFILES SHEET

DATE 3 May 1972

NUMBER OF EVALUATORS 5

PRODUCT AREA The Learner and the Learning Process



PRODUCT CODE

A _____

B _____

C _____

D _____

E _____

F _____

G _____

H _____

I _____

J _____

6-72-11

END