DOCUMENT RESUME

ED 065 541 TM 001 477

AUTHOR Johnson, Thomas J.

TITLE Evaluation in the Context of Product and Program Development in Laboratory and Research and Development Centers.

INSTITUTION Central Midwestern Regional Educational Lab., St. Ann, Mo.

SPONS AGENCY Office of Education (DHEW), Washington, D.C.

PUB DATE Apr 72

NOTE 121p.; Presented at AERA meeting (Chicago, Ill., April, 1972)

EDRS PRICE MF-$0.65 HC-$6.58

DESCRIPTORS Behavioral Sciences; Data Analysis; Decision Making; Educational Research; Evaluation Techniques; Federal Programs; Laboratories; *Models; Organization; Problem Solving; *Program Development; *Program Evaluation; *Research and Development Centers; Social Sciences

ABSTRACT

An effort is made to present a model of the R&D process which is comprehensive enough to handle the full range of events and activities which take place within it, including evaluation, and which has compatability at both macrostructural and microstructural levels of analyses. The first part of the paper focuses on the R&D process. An overview is presented of the federal R&D effort in education as it is accomplished through the labs and R&D centers and the scope of activities undertaken in one of these institutions. One of the chronic problems of evaluation, the degree of fidelity in the implementation of planned intervention is presented, and some procedures for dealing with the problem are explicated. The second part of the paper focuses on the products that emerge from the R&D process, a process conceived to be a mapping from the precise definition or characterization of a product at one stage into the definition of a product at a later stage. The minimal definitions or characterizations of products at the basic and applied research stage, the development stage, and the utilization stage are discussed for the physical sciences, and for the social and behavioral sciences by means of a series of three-dimensional minimal definition matrices. The paper concludes with a discussion of the evaluation of the federal R&D effort at the federal level. (Author/CK)

EVALUATION IN THE CONTEXT OF PRODUCT AND PROGRAM DEVELOPMENT IN LABORATORY AND RESEARCH AND DEVELOPMENT CENTERS

Thomas J. Johnson

National Program on Early Childhood Education

CEMREL, Inc.

Published under the National Program on Early Childhood Education of CEMREL, Inc., a private non-profit corporation supported in part as a regional educational laboratory by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

TABLE OF CONTENTS

INTRODUCTION

Although much has been written about aspects of the R&D process in universities and in labs and R&D centers, and much has been written about aspects of the evaluation of educational programs, and educational products, there still is a need to develop general models of the R&D process which are comprehensive enough to handle the full range of events and activities which take place within it, including evaluation and which have compatability at both macrostructural and microstructural levels of analyses. This paper represents an attempt to provide such a model.

The first part of the paper focuses on the R&D process. The paper presents an overview of the federal R&D effort in education as it is accomplished through the labs and R&D centers and the scope of activities undertaken in one of these institutions. The organizational structure to accomplish the work is presented and the organization for evaluation is discussed as it applies to this organization. The many decision makers and audiences for whom the evaluation effort can be directed, the political context, and the interplay and sequencing of these evaluation activities over time is discussed in some detail, and some important issues are drawn out.

One of the chronic problems of evaluation, the degree of fidelity in the implementation of a planned intervention is presented, and some procedures for dealing with the problem are explicated.

The second part of the paper focuses on the products that emerge from the R&D process. The R&D process is conceived to be a mapping from the precise definition or characterization of a product at one stage onto the definition of a product at a later stage. The minimal definitions or characterizations of products at the basic and applied research stage, the development stage, and the utilization stage are discussed for the physical sciences, and for the social and behavioral sciences by means of a series of three-dimensional minimal definition matrices. The evaluator's major activities are seen as focusing on (a) the adequacy of the mapping, or (b) the degree of specificity of information in the minimal definition matrices. Characterization of evaluation activities in terms of the data and the data and the data analysis is presented.

The paper concludes with a discussion of the evaluation of the federal R&D effort at the federal level.

PART I

THE PROCESS OF EVALUATION

There are at least four discrete types of activities which need to be completed before any major innovation gets adopted by the educational system of the nation. Although these four activities are sometimes labeled in different ways, we shall follow the usual procedure of referring to them as Research, Development, Diffusion, and Utilization. Each of these activities is distinguished by its form or purpose.

> Research--The primary purpose in undertaking Research activities is to augment existing knowledge about phenomena and their properties or to obtain new knowledge: to design and test models and theories concerning these phenomena, using a variety of analytical or empirical methodologies.

> Development--The primary purpose in undertaking Development activities is to use the observations or findings of research to design, engineer, produce, and refine products, i.e., instructional materials, procedures, and techniques which reflect and use this knowledge to attain desired outcomes. In some instances, a model or program will be developed which consists of a particular configuration of product components or elements of a system.

Diffusion--The primary purpose in undertaking Diffusion activities is to demonstrate products or models, and to inform, publicize, and interest the potential users of the products about their characteristics and features.

Utilization--The primary purpose in undertaking Utilization activities is to facilitate the adoption, installation, monitoring, and maintenance of the products and models as they are distributed to the user group.

Historically, the responsibilities for conducting these separate activities have seldom been allocated exclusively to any one type of institution or organization. The overall responsibility for research has most often been assigned to the universities and to some extent to private industry. The responsibility for educational development has been carried primarily by educational publishers and local school systems. The responsibility for the conduct of diffusion and installation activities has been covered by a multitude of organizations and agencies, including chiefly the state departments of education and the United States Office of Education, but also including schools, universities, and other educational agencies at the state and national level, and in some cases private industry.

Through the Elementary and Secondary Education Act (Public

Law 89-10), Congress set up a network of Regional Educational Laboratories and Research and Development Centers to help concentrate financial and human resources specifically on some of the research and development activities needed to help solve a number of the major problems in education. A hypothetical model or structure for conceptualizing the national R&D effort and the possible rationale underlying the genesis of the Educational Laboratory and R&D Center programs is shown in Figure 1.

Insert Figure 1 about here

The information required to rationally plan and implement national programs is derived from extant knowledge, technology, and data bases. The decision structure and the action programs serve to utilize and modify or to augment these various bases. The idealized sequence of activities moves linerarly from the assessment of national needs and identification of a target population (via a data base derived from census and other sources) to the examination and analysis of the knowledge base for possible solutions and the specification of needed technology. The current technology base is then assessed for the existence and availability of technology. If all stages of the sequence produce favorable information at this point, the design, implementation, and evaluation of remedial or preventive programs at the national level can be initiated. When the

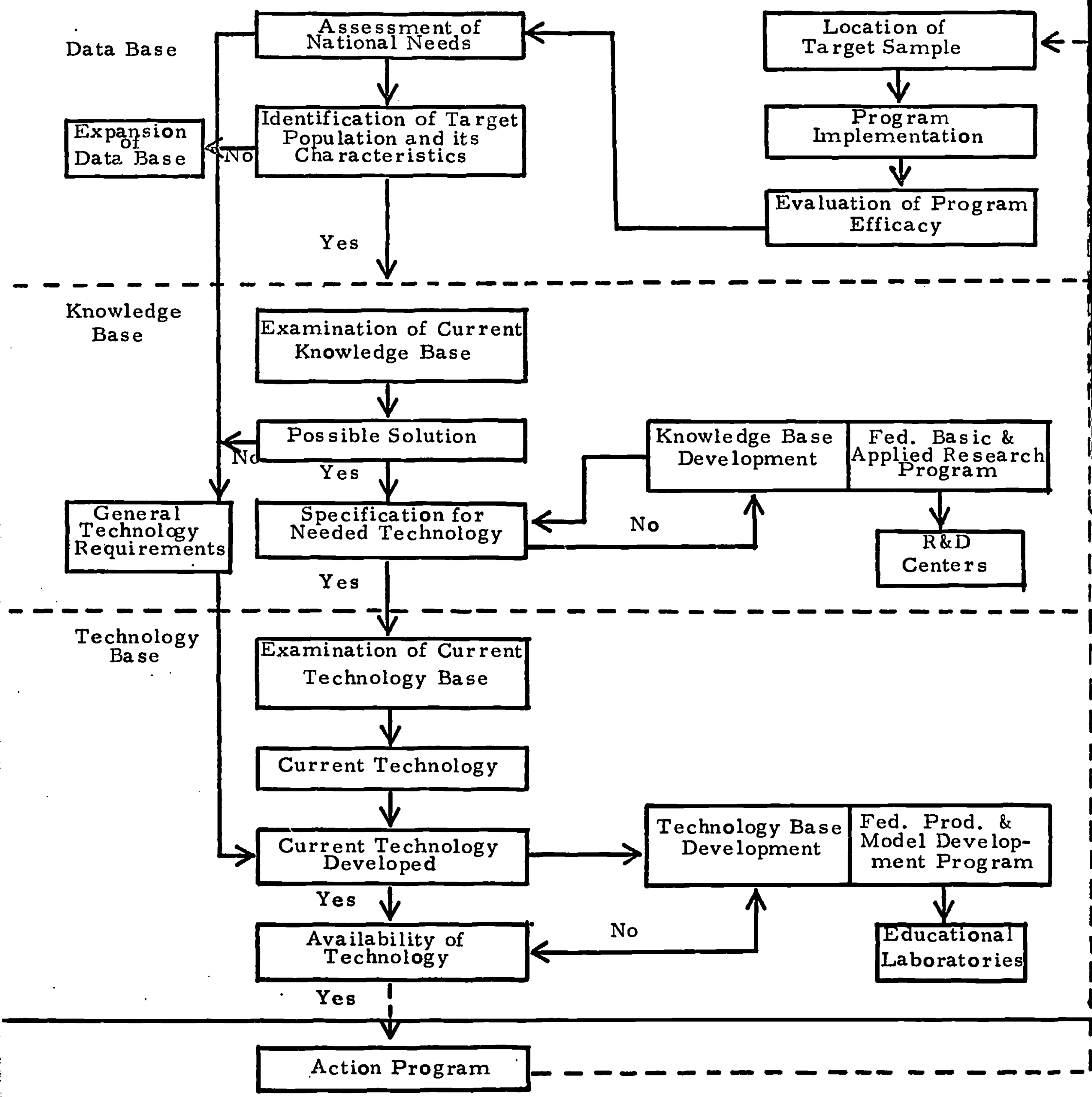


Figure 1

Hypothetical Decision Model for Federal R&D Programs in Education and Related Areas knowledge base is lacking necessary information, this provides the incentive for undertaking or sponsoring new research for knowledge base development and expansion. When the technology base is lacking, the general rationale is provided for undertaking or sponsoring programs and projects for the development of missing technology or technological systems. Sponsorship and support of institutions to achieve these outcomes play important roles in the total federal effort. Even, however, if the knowledge base does not provide the solution and specifications for needed technology, general solutions may be demanded by the urgency of the problem. This situation might, then, diminish the emphasis on basic and applied research supported by federal funds, since the rationale could not be based on its possible application to a manifest national problem, or to a technological requirement.

Operating within this framework, CEMREL, Inc. is one of twenty regional laboratories and research and development centers funded under Title IV of the ESEA Act. CEMREL is concentrating its energies and resources on the development of technology related to curricula and instructional systems in Mathematics, Aesthetics, and Learning Disabilities. CEMREL also has responsibility for the operation and governance of the National Program on Early Childhood Education (NPECE). In NPECE, both research and development activities are under way related to the knowledge and technology bases in the early childhood area.

Research and Development Functions

The scope of activities involved in the research and development process and the sequence of these activities is outlined in Figure 2. The activities that are primarily included under the research rubric include: (1) Reviews of research finding, theories, and methodologies, (2) theory design and revision, (3) exploratory and confirmatory research studies, and (4) design of investigatory methodologies. The activities that are primarily included under the development rubric include: (1) Reviews of existing technology, (2) product specification and design and production of prototypes, (3) design of evaluative methodologies, and (4) test and revision. Reviews of research findings and methodologies and confirmatory research activities may be legitimately included under the development label as well.

Insert Figure 2 about here

General Organizational Structure

In order to accomplish its program objectives and to complete these research and development activities, CEMREL has been organized into program and program-support groups, reflecting the functional relationship between the various activities. Figure 3 shows this relationship in an organizational function matrix, which has the separate research and development programs as columns, and the various functional R&D

Figure 2

Scope of Educational Research and Development Activities

Primarily Basic Research	Primarily Applied Research	Primarily Develop-ment	Type of Activities	Outcomes
		X	[a]Program design, specification of needed elements	Selection of Problem Solution
		X	Element design, specification of needed products, technology, desired outcomes, criteria	
X	X	X	Reviews, analyses, and summaries of research findings, theories and theoretical constructs and investigatory methodologies	Knowledge Base Development
X			Conceptualization, theory design and test	
X			Phenomenal description, characterization, exploratory studies	
	X	X	Confirmatory, verificational research	
X	X		Revision of theoretical models	
X	X		Design of investigatory methodologies	
X	X		Design of investigatory methodologies	Technology Base Development
	X	X	Explication of regulating variables, critical factors	

Primarily Basic Research	Primarily Applied Research	Primarily Development	Type of Activities	Outcomes
	X	X	Reviews, analyses and summaries of existing technology, materials, and requirements	Technology Base Development (continued)
	X	X	Product specification	
		X	Design of evaluative methodologies	
		X	Packaging design and prototype production	
	X	X	Product characterization	
		X	Initial test-revision cycle	
		X	Instructional system specification for product users	
		X	Pilot and field test and revision of product	
		X	Materials ready for spin-off	
		X	Configuration, integration of products of elements	Problem Solution Development
		X	Element packaging, prototype production	
	X	X	Element characterization	
	X	X	Test of elements and revision	

Primarily Basic Research	Primarily Applied Research	Primarily Develop-ment	Type of Activities	Outcomes
		X	Elements ready for use	Problem Solution Development (continued)
		X	Configuration, integration	
		X	Program packaging, prototype production of components	
		X	Program characterization	
		X	Pilot test and revision	
		X	Program ready for use	

[a]Program as used here refers to a configuration or system of individual elements, each of which contains one or more products.

activities performed by program staff or program-support staff on the rows. A more detailed analysis of the types of activities required in various phases of development is explicated in the Appendix. The four major CEMREL R&D programs are the Comprehensive School Mathematics Program (CSMP), the Aesthetic Education Program (AEP), the Instructional Systems Program (ISP), and the National Program on Early Childhood Education (NPECE). Program staff are assigned to functions which require background experience and technical expertise in the particular program area, while program-support staff are assigned functions which tend to be recurring across the separate programs. This is done to achieve greater efficiency and economy in operation, as well as to provide a wider work base for employing personnel with greater technical skill in these latter functions. There is considerable interplay

Insert Figure 3 about here

between the members of the evaluation staff and these other two groups. Thus, while the staff of one group may have formal or shared responsibility for an activity, the staff members of the other groups are consulted, or informed about the plans or accomplishments to which these activities are directed.

Organization for Evaluation

In order to complete its activities and responsibilities in an efficient

Figure 3

CEMREL Organization for R&D Activities
Organization and Function Matrix

	Laboratory Knowledge and Technology Base Development Programs			
	CSMP	AEP	ISP	NPECE
Program Staff Functions				
a. Review of research, theories, methodologies				
b. Exploratory and confirmatory experimentation				
c. Design of investigatory methodologies				
d. Reviews of existing technology				
e. Product specification and product design				
f. Product Development				
g. Product Revision				
h. User System Specification				
Program Support Staff Functions				
a. Administration and Program Management				
Administration				
Other Services				
b. Design of evaluative methodologies and product evaluation				
c. Communications				
Public Information Office				
d. Prototype Production				
Design Services				
Graphic Services				
Reproduction Services				
e. Program Diffusion				
Training				
Demonstration				

manner, the CEMREL evaluation staff is organized with a central staff of evaluators and an evaluation staff unit assigned to each program. This is shown in Figure 4.

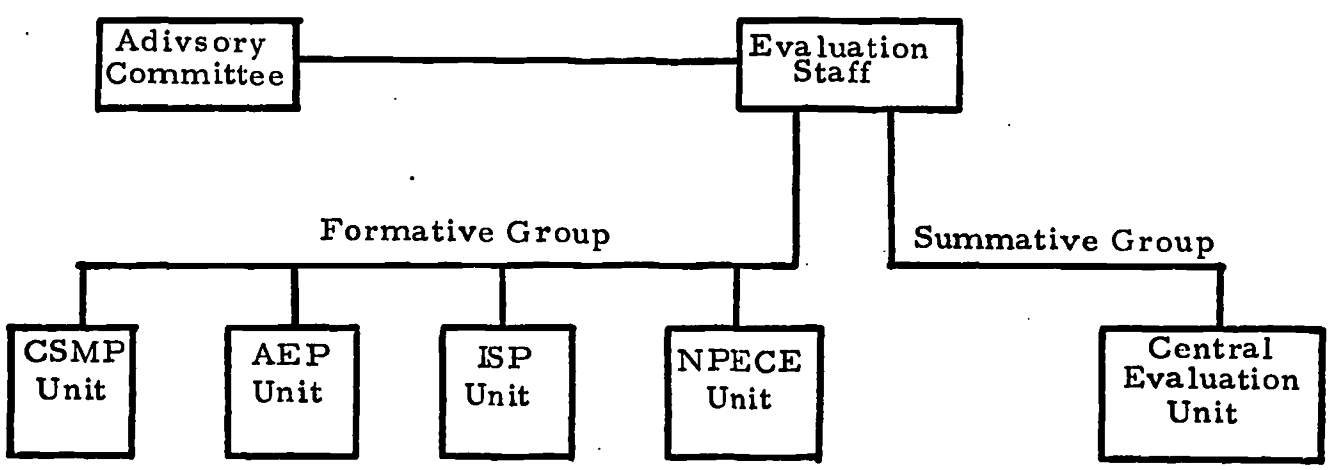
Insert Figure 4 about here

The separate program evaluation units perform primarily what Scriven (1967) has referred to as "formative evaluation" activities or roles, while the central evaluation staff performs summative type evaluation activities, and gives technical assistance to the program units when needed. One of the distinguishing features of the CEMREL evaluation organization is the existence of an Evaluation Advisory Committee, composed of nationally recognized evaluation specialists. This committee reviews the evaluation designs and strategies, discusses policy issues, and provides technical consulting on evaluation problems. The present members of the committee include: Michael Scriven, Berkeley; Gene Glass, Colorado; David Wiley, Chicago; and Al Beaton, ETS.

The previous discussion has attempted to explicate a macro-structural model for viewing educational research and development programs at the federal level, to explicate some of the activities that are required, and to show in fairly detailed manner how one R&D institution has been organized to perform this work at the program and program-support levels. The foregoing model of research and development and

Figure 4

CEMREL Organization for Evaluation



evaluation at CEMREL is intended to focus attention on the many environmental, political, social, organizational, and fiscal issues which influence evaluators, operating within a research and development institution. The organization for evaluation at CEMREL represents an attempt to deal with the constant problem of evaluator independence from program, accountability and responsibility issues, as well as other problems. The next section of this paper will provide a rationale for undertaking evaluation activities in laboratory and R&D centers and a conceptual framework for resolving some of these evaluation issues.

Scope of Evaluation Activities *

Decision Model

Evaluation activities may be described within the context of decision making. The process of making a decision may be thought of as consisting of four distinct phases.

1. Identification of the problem, and specifications for alternative choices, or solutions.
2. Specification of the various types of information which are necessary in order to have a basis for making the decision.
3. Design and formulation of appropriate means or modes of gathering the kinds of data which will be needed to provide the desired information.

---

*The conception of the major ideas presented in this section and the section following was done in collaboration with Martin Herbert and Louis Smith of the CEMREL staff.

4. Weighting of the information received against available alternatives.

Under the most general meaning of decision making, it is obvious that in making both trivial and basic decisions, many decision makers do not consciously conform to this process, especially in the early developmental stages of a product, nor should they necessarily do so. A simple example of this would be the decision to use a particular scheme of color coding to identify certain curricular materials. The information necessary might reasonably be restricted to the limited experience and opinions of certain members of the development team and the way to gather this "data" might be simply to arrive at a consensus at a meeting of those concerned. Unfortunately, many very important decisions are made in this rather informal manner, and the quality of some of these decisions might be improved if a stronger decision making process were operating. It is hoped that, by thoroughly investigating the wide variety of elements in this process, the interrelationships between the elements, the priorities, and the sequences in which decisions may take place, that the limited funds available for evaluation within any developmental project may be used more intelligently.

Dimensions of the Decision Model

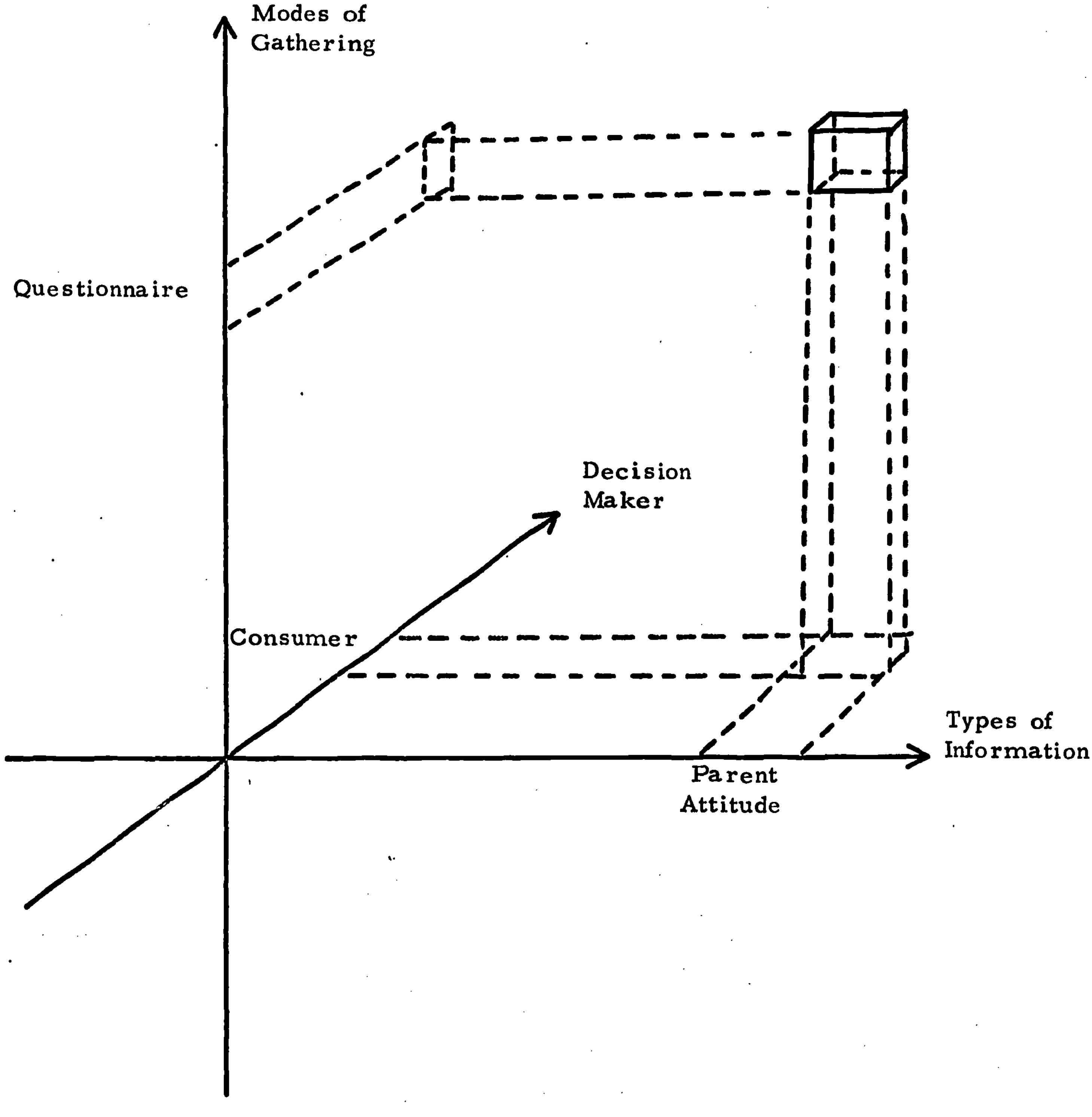
The role of the evaluation staff in the decision making process is that of gathering data to provide information to a decision maker. Each of these three categories, mode of data gathering, type of information, and decision maker, may be thought of as a dimension corresponding to a typology in a three dimensional decision making matrix. As a simple example, we will choose to take the consumer (a parent in this case) as the decision maker. Parent attitudes toward some aspect of the product might be the type of information desired, and a stratified random sample of subjects are to be given questionnaires as the mode of data gathering. This particular cell in the decision making matrix is shown in Figure 5.

Insert Figure 5 about here

Before describing in detail each of the dimensions, three brief comments can be made. First, as will be apparent when all elements along each dimension have been defined, not all cells of the matrix would necessarily be used, nor perhaps even be meaningful. Second, although one reason for conceptualizing the model in a three dimensional framework is to show the dependence of a mode of data gathering on the type of information and on the decision maker, the distinction between modes of data gathering and types of information may become blurred at times. Finally, the activities described within the matrix may not all fall under the exclusive domain of the evaluation staff. In fact, it is hoped that the model, as it is exemplified in a particular program, will help delineate the tasks to be performed, the kind of talent necessary to perform them,

Figure 5

Parent Attitude Cell of Decision Matrix



and the allocation of responsibility for their completion. One of our latent purposes is to break down some of the constraints that professional evaluators face in the information that they customarily deal with and its presentation format.

The Decision Makers

The decision makers who might be formally or informally concerned with evaluation results can be enumerated as follows. We have tried to characterize these groups according to their interest in the information and the use they will make of it.

1. The Developers include the administrative staff, the program advisory committees, production and technical staff, professional consultants to the programs, and teachers of pilot classes who work closely with program personnel. This group makes a greater number and variety of decisions and probably uses more evaluative information than all other decision makers combined. Most of what Scriven (1967) has called "formative evaluation" takes place in collaboration with the developer. This evaluation is focused primarily on the information the developer requires for decisions affecting continued technological development.

2. The Consumer. Potential consumers of education products vary from the individual teacher, department chairman, or

principal of a school, to the director or superintendent of an entire system, to a state department of education official. Decisions may be made by persons as disparate in their level of sophistication as the novice classroom teacher, the school psychologist, or state department of education officials. Unfortunately, much criticism has been heaped on these various groups because of the motives and the criteria they use in evaluating and selecting products for adoption and use; but it should be noted that the information they need is often not readily available, and when it is, it may be uninterpretable without detailed knowledge of the sources and contexts from which it was derived. Similarly, certain special techniques, such as those used in measuring cost effectiveness, are not available to them as part of their normal professional training; nor do they often get the basic statistical and experimental experience necessary for dealing with technically written evaluation reports. This is not meant to imply criticism of their professional training. Usually the consumer is not provided with a statement of the goals of a product in a way which allows him to distinguish it, at a proper level of generalization, from potential competitors. He is often not provided with an accurate and faithful characterization of the product, and he may not be presented with information which allows him to judge the adequacy of the product in terms of his own educational goals, especially where these differ from the avowed goals of the developer. Thus, it is not surprising that the explanation given by the consumer for a given decision may seem inadequate.

3. The Sponsor includes federal, state, and local education agencies, private foundations, and private industry. The decision to fund or to continue to fund is usually based on less specific information than the consumer needs. Of primary interest to the sponsor is the investigation of the developer's plans to evaluate and meet the needs of the consumer. The sponsor is also interested in the scheduling of development and diffusion strategies; not so much the details of production, as the developer's plans for meeting various contingencies in the process used to attain the product. Although the sponsoring agency will make its own plans for obtaining whatever information it thinks necessary, it is incumbent upon the developer to ascertain what information is or will become necessary, in what sequence, and at what level of development, and to incorporate into his plans specific methods of gaining and transmitting this data so that he is prepared to provide it as needed.

4. <u>The Skeptics</u> or potential critics may be divided into at least three partially independent subgroups. One group of skeptics are those who are interested in substantive issues. Included in this group are the subject matter specialists, whether the subject is an academic discipline or an instructional technique. The evaluative criteria used by this group are, generally, intrinsic in nature, based upon the specialist's own knowledge and experience and upon the congruence of the product to some well established conceptual framework he employs. A second group of potential critics are the methodologists. This group partly encompasses the peer group to which the evaluation staff belong (i. e., the research community). The foci of their interest are the actual evaluation procedures used and the relevance of the data yielded to describing the product's adequacy. A third subgroup may be identified as the practitioners, who may be asked or coerced into accepting a new product. These individuals belong to both the consumer group and the skeptic group, but their influence may be a factor in development before the stage of potential adoption. Their criticism is often directed toward the practicality of introducing and maintaining the new product, given the constraints they perceive. It is

obviously a rather simplistic view that potential critics fall into such mutually exclusive categories as intrinsic, empirical, and practical, since any critic must at least touch on part of each category. Nevertheless, it is useful to clarify criticisms of a product into these three areas, since each dictates possible evaluation strategies that might be implemented.

5. True Believers are diametrically opposed in viewpoint to potential critics. They have accepted the product (or the philosophy behind it). Members of this group usually base their opinions on one of the three criteria mentioned previously, i. e., implicit evaluation based on rational analysis of the product, empirical evaluation based on measurable evidence of the achievement of the aims of the product, or a belief that the product is amenable to practicalities of the real world and will be liked by the users. Thus, there is a certain parallel between the skeptics and the true believers. Both usually confine their interest to one main category of evaluative criteria; the skeptic for a source of investigation and potential criticism and the true believer for a source of sustenance for his belief in the face of criticism on other grounds.

6. <u>Legitimators</u> are persons whose professional stature gives weight to their expressed opinions. They provide opinions or evaluations of the product, which the educational community will accept without further personal investigation of the information. Advisory Committees, National Committees, boards of directors, distinguished visitors, and consultants, may be placed in this category.

7. <u>Other Developers and Cooperating Institutions</u>. Laboratories, research and development centers, and cooperating universities comprise the final group of decision makers. They provide a pool of experience in dealing with funding agencies, developmental procedures, staffing requirements, diffusion strategies, and research information. They provide a pool of talent both directly, through staff, and indirectly, through the advisors and influential people with whom they associate outside their institutional context. Insofar as continued cooperation is necessary to maintain adequate staffing, the supply of information to these institutions must be maintained.

For each of these groups of decision makers, different types of information must be provided, and for each type of information for a given decision maker there may be various methods of obtaining that information. These methods will be discussed in the next section. Following that explanation, we will consider the associated problems of priority and sequencing of evaluative tasks. With limited funds, staff, and time available, all tasks cannot be performed, and the sequence of performing even the possible tasks has certain constraints. The nature of these constraints and the priority of possible tasks will be dealt with in a later section.

Types of Information

Information about the product will usually include three essential aspects. How can the product be described or characterized (product specifications)? What educationally relevant outcomes can be expected from it (performance specifications)? What sacrifices must the user make to adopt the product (systems specifications)? The description of these types of information will be very brief here, since most items are self-explanatory.

Product Specification

1. The target population, in general, refers to the final users, usually students. Age, stage of cognitive development or level of academic achievement, behavioral characteristics, and socioeconomic factors may be some of the useful descriptors for characterizing the user group, and limiting the tested applicability of the product.

2. Duration of the use of the product both for an individual user (usually a student) and for the whole adopting agency.
3. Content, (a) explication of the multivariate set of major and minor independent variables that are assumed or known to be regulated or "delivered" by the product, or (b) the structure of the knowledge it purports to convey.
4. Delivery system, including media.
5. Procedures and tasks to be carried out by the implementor in mediating 3 and 4.
6. Goals. This item is used in a dual sense. The final product should have goals at various levels from rather vague, general, educational goals through intermediate stages to rather well-defined, behavioral goals. All decision makers need this information. But the process through which goals become adopted by a developer, the refinement and revision of them as the product emerges, and the decisions made by the developer about the relative priority of various goals are crucial and will recur throughout this paper. These issues are inextricably woven with various performance specifications as well.

Performance Specifications

7. Explication of the multivariate set of student or user outcomes.

8. Student or user attitudes.

9. Teacher and administrative attitudes.

10. Parental and community attitudes.

11. Interactions (if any) between user and product.

12. Student capabilities in a wider sense than is usually connoted by "student achievement." An example of this might occur, for example, in the explication of a general psychological outcome accompanying the use of a particular product series or sequence over time.

Systems Specifications

13. Financial costs.

14. Changes in organizational structure and personnel tasks required to use the product.

15. Training requirements preservice and inservice.

16. Physical equipment requirements.

17. Duration of commitment.

18. Monitoring requirements.

The definition and interaction of items 3 and 7 and the choice of modes of data generation to assess this interaction remains the central problem in product and program evaluation. At a minimum, documentation on the assumptive, logical, or empirical relationship between the major elements of specification for the product and the major performance specifications should be provided. The conventional modes of data generation are listed below. Factors related to the selection of one or more of these instruments will be dealt with in the final section of this paper.

1. Tests, instrumentation
2. Devices, apparatus
3. Conventions, rules
4. Calculi, formulae
5. Simulations, representations
6. Naturalistic observation
7. Search through relevant literature, research findings, other information bases
8. Expert opinion
9. Philosophical analysis

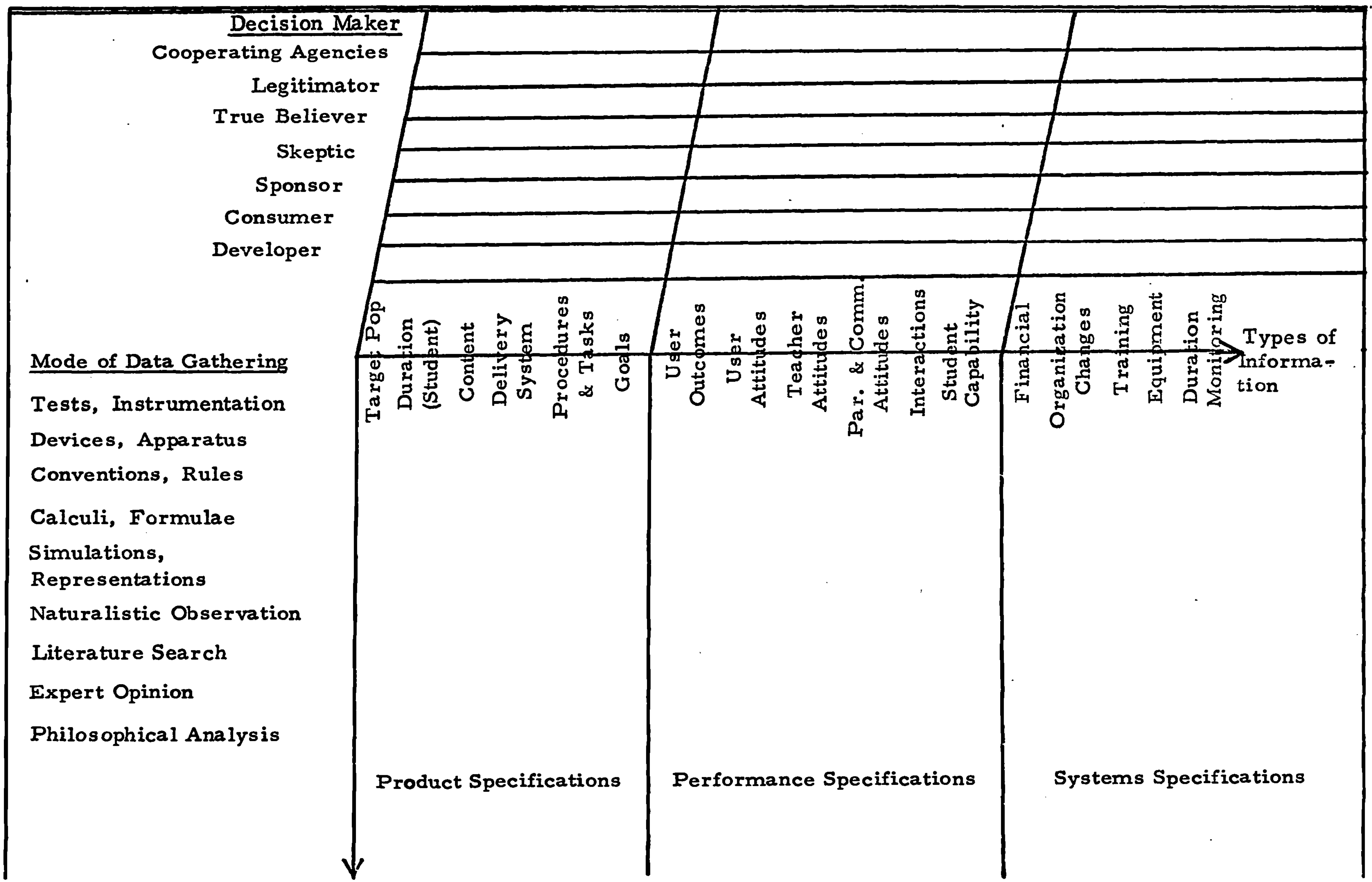
The overall structure of the decision matrix is depicted in Figure 6. All evaluation activities undertaken at CEMREL on behalf of its

Insert Figure 6 about here

products can be characterized by their location in one or more of the cells of this matrix. Evaluation activities can, we believe, have wider value and impact when set in the context of the decision matrix, because the particular kinds of information and the audiences addressed are clearly identified. The matrix thus defines the role of the evaluation unit. It

Figure 6

The Decision Matrix



should be alert to the information required by different decision makers and should provide appropriate data in as many cells of the matrix as possible over the development period.

The Evaluation Process Over Time

Both practical and theoretical problems arise when the evaluation process is considered from a time perspective. The most important of these are the managing and sequencing of evaluation activities. While each part of the model might be elaborated in this context, we will isolate only two or three illustrative issues.

Technology and Product Development as a Process

Sponsoring and funding agencies such as the U. S. Office of Education presumably hold conceptions and viewpoints regarding the organizational and intellectual problems in developing new technology. One possible conception might involve a series of events which we have described over simply in Figure 7.[2] Such a cycle might run over several

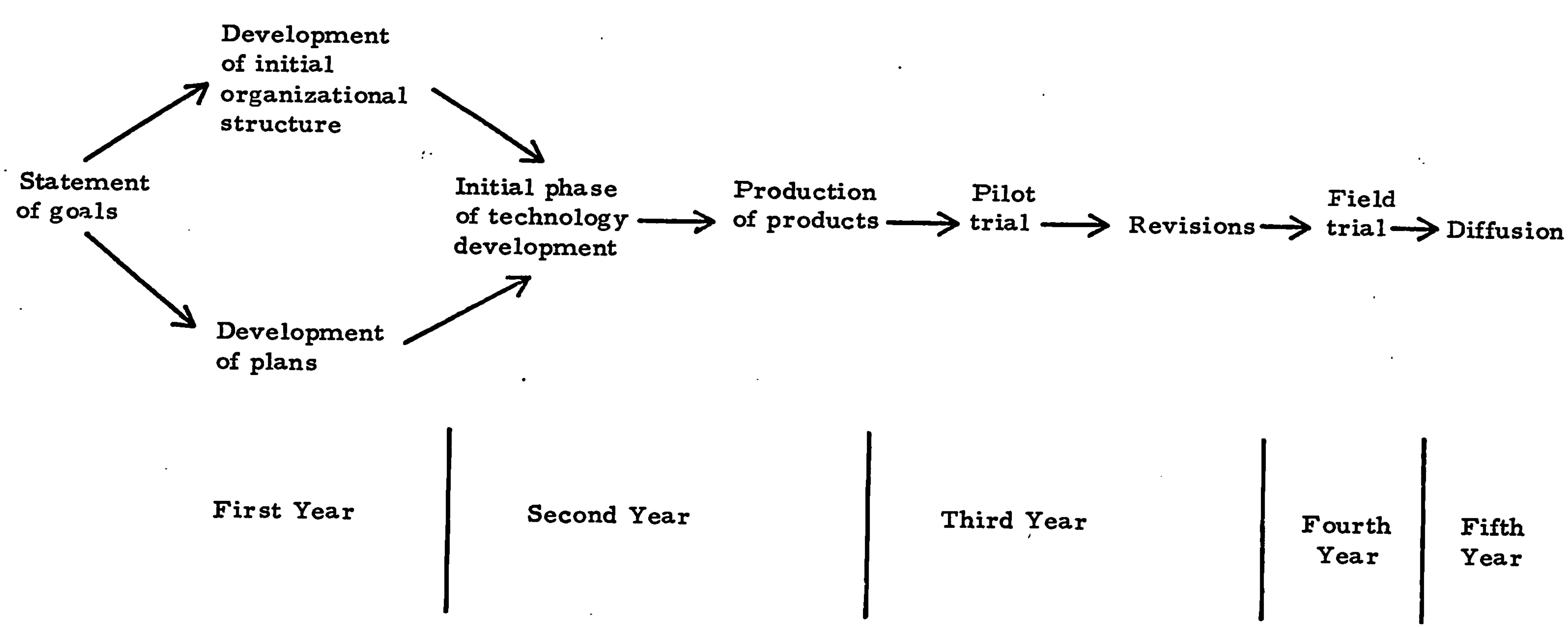
Insert Figure 7 about here

years and be complicated by several strands of activity, e.g., separate groups and individuals working at different grade levels or on different

[2]More elaborate statements for each program are found in CEMREL's five year planning documents.

Figure 7

Idealized Process of Technology Development



topics within a broadly conceived new curriculum, incorporating the new products.

For example, the first year's activities might involve planning and development of organizational structure, procedures, hiring, and training of staff. The second year might see prototype materials developed early and later usable classroom products. The third year might include a pilot trial in several classes, plus revisions. The fourth year might involve a large scale field trial with varied children, teachers, schools, and communities. And the fifth year might be spent on regional or national diffusion. Such a scheme is stylized and rationalized as an ideal type. The actual process involves considerable interplay among parts, revisions, updating and "schedule slipping," as any educational developer will realize. However, such complications, while real and in need of discussion elsewhere, are only complexities of the basic issue in the present discussion.

At each point in the process, different evaluation activities and skills--and probably personnel--are required if the demands made by the funding agency are to be met. A funding agency review team, which has read and accepted the work of Scriven (1967) and others, might want an evaluation of the program goals. The essential question concerns the worth of what a new development program is attempting to do. The analytical skills required and the tasks involved in this very simple but basic question have not been well answered by scholars in the evaluation community at large. Presumably, some combination of sophistication in the content field and in philosophical and ethical analyses would be required. If this is, in part, a political question, i.e., What does the community want for its children?, then both political theory and associated empirical investigatory skills are required. For example, if a new math program such as CSMP or CAI can teach some defined mathematical "common essentials" in twenty minutes/day rather than a usual forty minutes/day, we have an interesting decision problem. The "saved" twenty minutes accumulates to 100 minutes a week, a little over an hour and a half. Over a full year (180) days, this amounts to some 60 hours of instructional time. Do we spend the 60 hours "accelerating" students in math, "enriching" their math experience, or perhaps even in reducing the time spent in school? Or, is the question no longer a mathematician's problem but a political problem, in the best sense of the term political. That is, I, as one of the majority of parents, campaign, through the school board, for the introduction of a 60 hour program in aesthetic education; or, for a compromise giving teachers or the parents and their children a choice of more math, aesthetic education, or released hours in the saved time.

The point we want to make is that, in our model, "goals" is an important problem impinging upon the evaluation unit. If the goals of the new technology, or products, or the curriculum they are to be incorporated into are judged worthy upon analytical investigation, this is important information to the review team from the sponsoring agency. If the evaluation unit finds that most parents would like a combined enrichment and acceleration option in mathematics, a case is made for both the review team and for the director of the curriculum process to utilize, divert, or increase resources for developing these parts of the curriculum.[3]

To continue our illustration of information needed by our hypothetical review team from the sponsoring agency, after the goals have been set, and probably while the initial materials are being developed in the second year, and certainly before any extensive piloting during the third year, another simple and basic question arises: "Will the kids learn any mathematics in the new program?" The evaluation team must then convert goals and plans to specification tables and these to test items and these to scales with known validity and reliability assessments. A technology or product development program whose evaluation unit is not producing such "devices" during that period should be in trouble with its funding agency.

The Impact of the Product on Potential Adopters

Perhaps the most critical information desired by a consumer or

---

[3]Of course, other reasons and other data may also be critical in the decisions regarding resource allocation.

potential adopter is the amount and kind of pupil achievement outcomes that will be mediated through the new curriculum. The adopter's needs can be addressed by a mode of data generation relatively unique to the CEMREL evaluation model. This mode of inquiry we have called "naturalistic observation," "microethnography," or "participant observation." Our most intensive utilization of this technique occurred in our CAI evaluation (Russell et al., 1969). The problems at issue were two-fold: (1) the impact upon the organization adopting the curriculum, and (2) the impact of the curriculum upon the teacher who must carry out the instruction. Such data are seldom presented in careful detail to the adopter.

In the ethnographic part of the CAI report Smith and Pohland (1969), described CAI as a complex social and technical system. The problems involved in the utilization of the program in a mostly rural and mostly economically-depressed region are elaborated in detail. The cultural milieu posed its own set of contingencies. The funding of the program, which was essentially through Title III of ESEA, involved delays, cuts, and changes, which created severe problems for the administrators of the program. Similarly, personnel qualifications, technical difficulties in the transmission and servicing of equipment, and the research and development goals of the organization providing the program, all became important events in the world of the adopter. By carefully observing the pilot trials, describing the "realities" of the program, and conceptualizing the issues in more general terms, we carried out a phase of evaluation important to administrators who will have responsibilities for implementing such a new curriculum or instructional program.

The second major finding dealt with teacher and pupil behavior within the context of the program. Observations produced important data on "usual" teacher behavior, pupil involvement with the program, and such social dimensions as "competition" in an individualized program. While this report did not pursue the discussion to the point of development of a "theory of instruction" or a "configuration of independent variables" underlying the program, this has become a major question under investigation in our current efforts in the observation of CSMP pilot trials.

The kinds of knowledges, orientations, and skills demanded for this part of the evaluation are quite different from that demanded in other parts, such as goal analysis or test construction. Some sophistication in organizational theory and instructional theory is necessary to analyze and understand the issues and the concerns of school administrators and teachers. The initial development and utilization of the CEMREL evaluation model (Russell et al., 1969) illustrates the problems and potential of this viewpoint. In terms of the phasing of such activities, the most logical time for such an emphasis on naturalistic observation is during the pilot or field trials. Insofar as such data are important for the curriculum developer, the evaluation activity must come considerably earlier.

Interdependencies in Evaluation Activities

While a number of examples could be used to illustrate the interdependency of evaluation activities, we will use one which combines CEMREL's CAI and CSMP experiences. Observers of the CSMP program noted an occasional use of workbook-type drill sheets in classrooms. Presumably, the teachers were trying to help the children develop immediate recall of particular combinations or "facts." This kind of math objective received little emphasis in the CSMP program, and this kind of "lesson" was not contained in the activities which composed the packages. From this observation and conceptualization, a suggestion was made for the expanded pilot or field trials. The trials would have several conditions: regular individualized CSMP, supervised CSMP, teacher instruction with CSMP materials, and traditional math taught in the expanded pilot. In addition the concern for some drill activities, made evident from our naturalistic evaluation, and the knowledge of the high motivational qualities of CAI drill and practice routines made the possibility of introducing a CAI condition into later field experiment seem highly desirable.

While there are many ways in which such an experiment could have evolved, it happens to have occurred in the above fashion. In the same way, the format of the CAI evaluation, e. g., the feedback of time and error scores as they contribute to competition and motivation, awaits further experimentation. As in many of our illustrations, the data collected in these trials could be useful to several kinds of decision makers--developers, legitimators, adopters, and curriculum and instructional theorists who constitute reference groups relevant to various personnel in the program.

The management of an evaluation unit must be aware of such interdependencies and contingencies among projects, personnel, and tasks within the scope of evaluation.

Summary

The evaluation process over time refers to the issues of sequencing and managing the flow of activities and information developed by the evaluation unit. The conclusions we would offer include:

1. Evaluations must be cognizant of and contingent upon the context, e.g., the curriculum development per se.
2. Some parts of the evaluation must begin long before the utilization of the knowledge takes place.
3. The evaluation unit will receive different emphases and have different demands placed on it at different points in the total process of curriculum development.
4. Information gathered at one point in time may be critical for a different group of decision makers at another point in time.

5. Information generated by one orientation in one phase of evaluation may be a prerequisite for another part of the evaluation in a later phase.

6. Quick and dirty methods may be needed to "cover the bet." Carefully controlled experimentation may be a luxury to be engaged in only where the level of certainty and the consequences of wrong decisions require it.

Some Newer Developments in the Methodology of Program and Product Evaluation *

We have attempted to present information concerning the structure of R&D activities as they are under way in institutions like labs and R&D centers. We have taken one institution, CEMREL, and have attempted to show how the staff is organized to accomplish program activities and program-support activities, including evaluation. We have presented a model for conceiving of these evaluation activities in the context of multiple decision makers, who require various types of information. We have also tried to discuss some issues involved in the implementation of the evaluation process over the development period. Up to this point, there has been no discussion of any of the usual design and methodology topics that warm the hearts of evaluators; although we might mention in passing that we have found fractional factorial and response surface designs to be of

*The conception of the major ideas presented in this section was done in collaboration with William Wright and Dan Ferritor of the CEMREL staff.

real value in arriving at efficient evaluation strategies. In this section we will present a brief discussion of some of the less exotic methodologies that CEMREL has found to be useful in dealing with one particularly critical evaluation problem: the measurement of the degree of implementation of an instructional program.

Degree of Implementation

In our view, the measurement of the degree of implementation of an instructional program is one of the most critical problems facing evaluation personnel, and yet it has received far less treatment in the evaluation literature than its importance warrants. The problem arises out of a common type of technology development program, for example, where the developer has created a set of materials for training teachers to use a particular instructional strategy or technique in their classrooms. The most important information to collect for the evaluation of the efficacy of the training materials is not the performance of the children, but the actual degree to which the particular strategy is used by the teacher in her classroom. This is not meant to imply that the first is not an important set of data, but rather that the question of implementation is prior to the determination of any consequences presumed to be derived from it. Moreover, the "effectiveness" of the techniques in terms of child outcomes are presumed to have been manifest in the "hot house" trials, or in applied experimental settings.

For example, at the present time the staff of the Instructional Systems Program is developing a behavior modification, teacher training program. This program is an outgrowth of several years' experimentation at inner-city elementary schools in St. Louis and in other settings. The details of the intended curriculum may be found in the Basic Program Plan of the Instructional Systems Program.

The CEMREL Evaluation Unit has undertaken two basic activities in connection with the development of these curriculum packages. We have been observing some classrooms of teachers trained in the use of contingency management techniques by members of the Instructional Systems Program. We have also been developing behavioral objectives and test items from which we will construct paper and pencil tests to accompany the training units.

These efforts are designed to produce a three pronged approach to the evaluation of the General Reinforcement Packages. During the training sessions to be given to participating teachers, the tests developed by the CEMREL Evaluation Unit will be administered. The tests will be used again at the conclusion of the school year. The data from these tests should provide us with information concerning the degree to which the teachers have mastered the information presented in the packages, and their retention of that information.

Cognitive mastery of these packages is not a sufficient outcome, however. The important issue is whether or not the teachers implement the recommended procedures for classroom management. If, however, a particular teacher or group of teachers manifests a low degree of implementation, we will want to know whether that occurrence is a function of lack of understanding of what they are expected to do, traceable to some failing in the training program, or involves some decision on their part not to become involved in the process.

The degree of implementation variable is essentially the matching of classroom procedures used by the teachers with those intended by the program developers. We propose to attempt to measure the degree of implementation by classroom observations of the teachers' use of reinforcement. There are a host of measurement problems associated with such an attempt. For one thing, prior to any observation, it will be necessary to specify the domain of those behaviors which are expected, and, moreover, to establish the priorities of these if weighting seems appropriate. Given a particular type of classroom difficulty, for example, the most appropriate behavior from the program's perspective might be the institution of a Premack system, but if the teacher uses a modeling approach that might also be considered implementation, though at a lower level.

In addition, even given random assignment of teachers to experimental and control groups it may well be that there will exist gross differences in the frequency and type of behavioral problems that exist in the classrooms. Establishing the degree of implementation, given different levels of disturbance, presents another array of measurement problems.

There are also analysis difficulties associated with such an approach. One goal of the General Reinforcement Program is the alteration of student behavior, and better managed classrooms. We will, therefore, observe classrooms not only with a view toward recording teacher behavior, but also for the purpose of collecting data on the behavior of the students in these classrooms. Degree of implementation, then, may be classified as a dependent, independent, or mediating variable in different senses.

No doubt other issues will arise if we proceed with this approach. The objectivity and the reliability of the observations, for instance, will be matters of concern. The point is that we believe that this task, although complex and problematic, is important and desirable.

It should be made clear that this effort would be in a very real sense a feasibility study of the evaluation strategy as much as it would be the crux of the evaluation of the program itself. At a minimum, moreover, we would anticipate the need for a full-time staff member at the doctoral level and with a background in field observation as well as the contracted services of college graduates in the pilot site locale.

If we are to proceed with the evaluation as outlined above, it will be necessary to develop an observational system for use in the fall. The Instructional Systems Program staff has begun the task of identifying the critical teacher behaviors. In addition to the completion of the list of teacher behaviors, it will be necessary to specify the student behaviors to be observed, and to develop a record keeping system. No doubt the suggested approach to the evaluation of the General Reinforcement Program will be an expensive enterprise. There are alternative courses of action, but any evaluation would have to include classroom observation of teacher activities and student behavior.

The preceding sections of this paper have sought to present a model for conceptualizing a wider range of evaluation activities as they operate within an institution such as a laboratory or R&D center. A number of conclusions can be drawn.

1. Because multiple decision makers, other than the developer, require information about technology at an early stage, the specific distinctions between formative and summative evaluation that Scriven (1967) noted are likely to be blurred.
2. Because the resources allocated to evaluation are limited, the evaluation staff may be required to provide an array of information of potential use to each of these decision makers out of every separate evaluation activity it undertakes.
3. Because the evaluation activities do provide information to

these multiple decision makers, they all require a high degree of rigor and careful planning if they are to be valid and useful. In other words, formative evaluation is not a synonym for informal evaluation.

4. On many occasions the evaluation staff will be required to invent new and ad hoc methodologies to answer special questions of concern. Moreover, research in the methodology of evaluation, almost of necessity, precedes any research advances made in the given discipline or content area (e.g., aesthetic education).

5. Both intended and unintended consequences of the newly developed technology are reasonable concerns of the evaluator and require that he be independent of the developer.

6. During the early stages of development in limited settings, it might be feasible to use a variety of modes of data gathering, including non-empirical types such as non-participant observation. Evaluators, who are carefully trained in these techniques, may provide information of a unique character and of considerable value especially to the developer.

7. Because the evaluation staff plays a critical role throughout the various phases of technology development, the organization for evaluation should be sensitive to issues of independence,

bias and co-option by developers, and to reviews of its own plans and actions. The basic question as Scriven (1971) has so neatly phrased it is "who evaluates the evaluators?"

PART II

THE PRODUCTS OF EDUCATIONAL RESEARCH AND DEVELOPMENT *

Our experiences with the ISP Program characterization problem serves as a convenient point of departure from the general tenor of the previous sections of this paper. Up to this point we have presented laboratory and R&D center evaluation in terms of (a) the organizational arrangements for it, (b) the political context in which its activities are placed, (c) the range of activities and types of outcomes it encompasses, and (d) some vexatious problems that confront evaluators somewhat chronically.

In short, we have elaborated on the process of research and development and the place of evaluation in it. To some, the presentation may not serve to provide anything substantially new, or might be viewed as a restatement of the concerns that many other evaluators have thought, written, and talked about from time to time. However, missing from these discussions of evaluation (and ours up to this point) has been any focus on

---

*The author is indebted to Harry Kelly, Dan Ferritor, Robert Hess of the CEMREL staff, and to David Wiley of the Evaluation Advisory Committee for ideas which led to the formulation of the second part of the paper.

the product side of the research and development enterprise. This might seem to be a casual problem at first glance, but we are now convinced that specification of the form of the products of R&D as they move through the R&D process is the crux of the solution to well-planned, well-executed R&D in education and that failure to provide this specification is responsible for much of the non-lineraity of the R&D process that is commonly observed, and the source of many of the evaluation problems discussed earlier.

Special Factors in the Presentation

In the next part of the paper we will try to present a general conceptual model of Educational Research and Development from the product side which we have found useful in planning the National Program on Early Childhood Education at CEMREL, Inc. We will then attempt to show how critical it is to incorporate the evaluation process into this. The general R&D model is a complex one since it needs to deal with all of the R&D activities that might be under way, and the scope of the presentation will be on the total complexity. Before proceeding, we need (1) a clear understanding of the notion of mapping from elementary algebra, (2) a means of illustration of the model, (3) a system of notation for referencing illustrations and their component elements, and (4) a set of elements and their definitions where necessary.

The Concept of Mapping

At its simplest level, mapping is concerned with setting up the rules for connecting elements of any one set of things onto the elements of another set or onto itself.

A mapping (or function, or transformation) from A to B is an assignment which assigns to each element X within A ($X \in A$) exactly one element Y within B ($Y \in B$). The crucial aspect of the notion of mapping in that the "exactly one" relationship is necessary. The arrow diagram below in Figure 7 shows one possible mapping of the 5-element set $A = \{a, b, c, d, e\}$ to the 3-element set $B = \{j, k, l\}$.

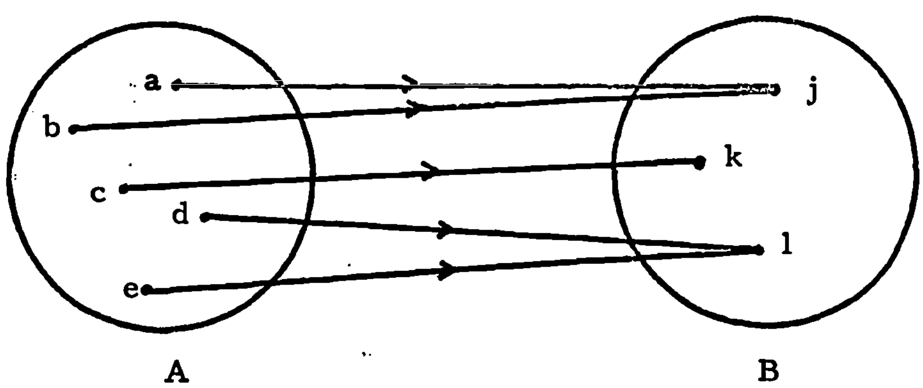


Figure 7

Note that for each element of the set A, there was an assignment to only one element in B. Thus, to the element $a \in A$ the function f assigns the element $j \in B$. When a mapping f from A to B assigns the element $y \in B$

to the element $x \in A$ we call y the f-image of x. Thus, in the function illustrated in Figure 7, the f-image of a is j, the f-image of b is j also, the f-image of c is k, etc., etc., and it can be noted that each element of A has only one f-image. We denote the f-image of an element $x \in A$ by the xymbol "f(x)," read as "the f-image of x," "the image of x under f," or "f of x." Thus, referring back to Figure 7, the f-image of a is j and this can be written as $f(a) = j$.

A classification system for mappings or functions, or transformations, as they are alternately called, can be developed based on the relationships between any two sets A and B not necessarily distinct.

1. A mapping f from A to B is a one-to-one mapping if and only if different elements of A always have different f-images in B.
2. A mapping f from A onto B occurs if every element of B is the image of at least one element of A.

Illustrating the Model

For illustration purposes, the model requires a configuration of three-dimensional matrices as illustrated below in Figure 8.*

*The model actually requires an n-dimensional space for explication. This will be done through a series of 3-dimensional projections. The dimensions may be binary, multichotomous, continuous, or probabilistic.

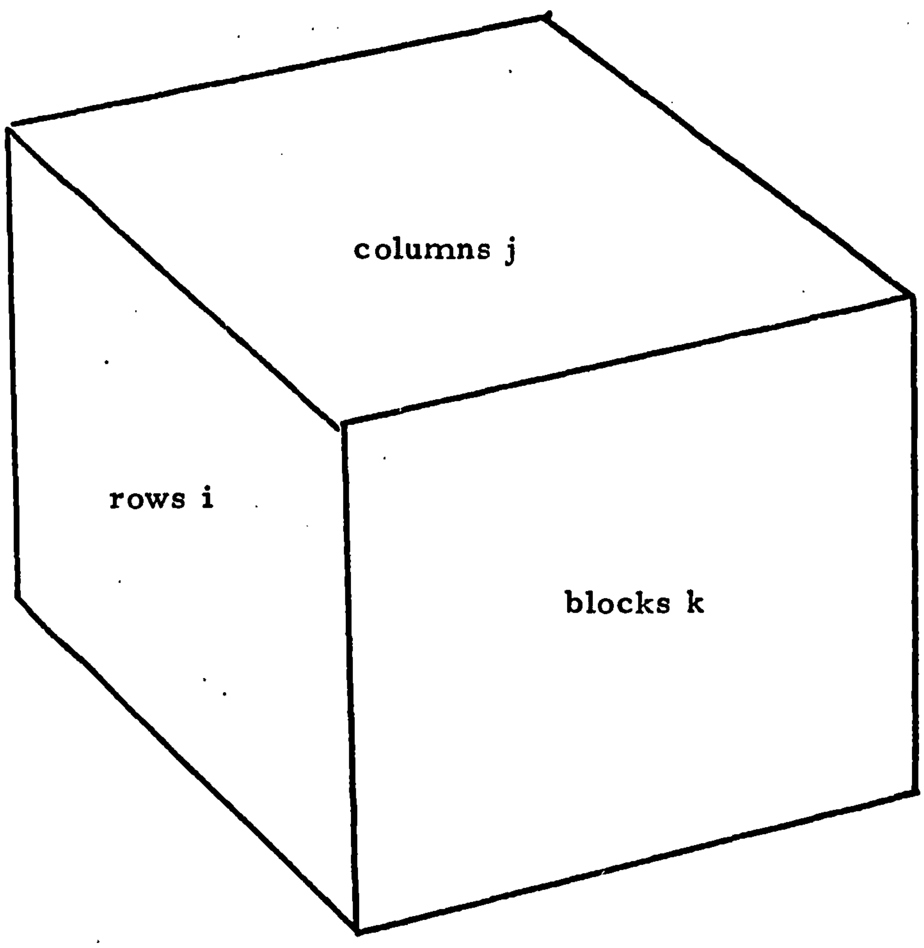
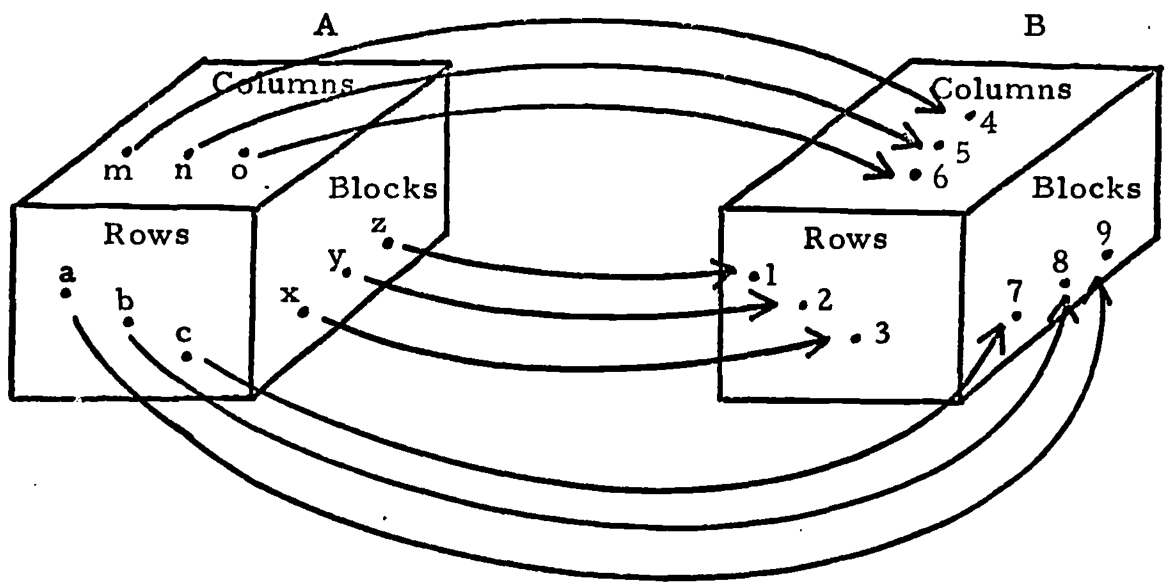


Figure 8

Basic Three-Dimensional Matrix

We will use the conventional notation and refer to the side facing the left of the page as the rows, i, of the matrix, the side facing the top as the columns, j, of the matrix, and the side facing the reader as the blocks, k, of the matrix. Thus, if we include the number of the figure as part of an identifying label, 8_i would be the rows of the matrix in Figure 8, 8_j would be the rows of the same matrix, and 8_k would be the blocks.

If the elements of a three-dimensional matrix A were to be mapped onto the elements of a three-dimensional matrix B, the figure illustrating this "flow" from A to B could be drawn as shown below.

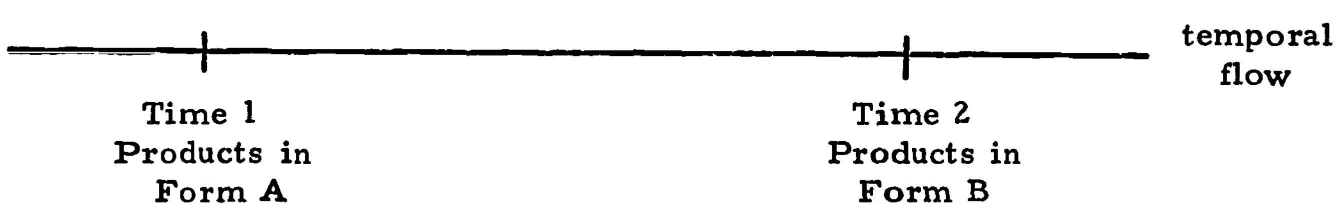


In the figure, the elements a, b, c, on the Rows of A have been mapped onto elements 7, 8, 9 on the Blocks of B. The elements m, n, o on the Columns of A have been mapped onto the elements 4, 5, 6 on the Columns of B, and the elements x, y, z on the Blocks of A have been mapped onto the element 1, 2, 3 on the Rows of B. Thus, as a general rule for our purposes, <u>any row, column, or block and its elements can be mapped onto any other row, column, or block and its elements</u>.

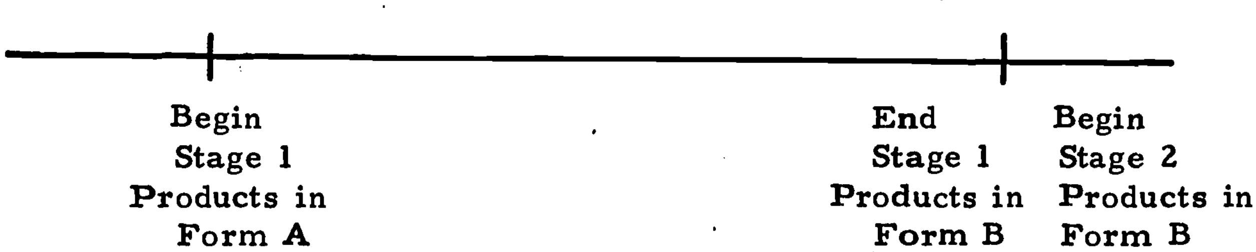
Note that the flow from A to B need not be only from row to row, column to column, or block to block. For example, x, y, z on the blocks of A is mapped onto the 1, 2, 3 of the rows of B.

While mapping has a number of uses in mathematics, the concept of mapping is especially useful in the illustration of process phenomena, and, thus, can be directly applied to the illustration of the R&D process.

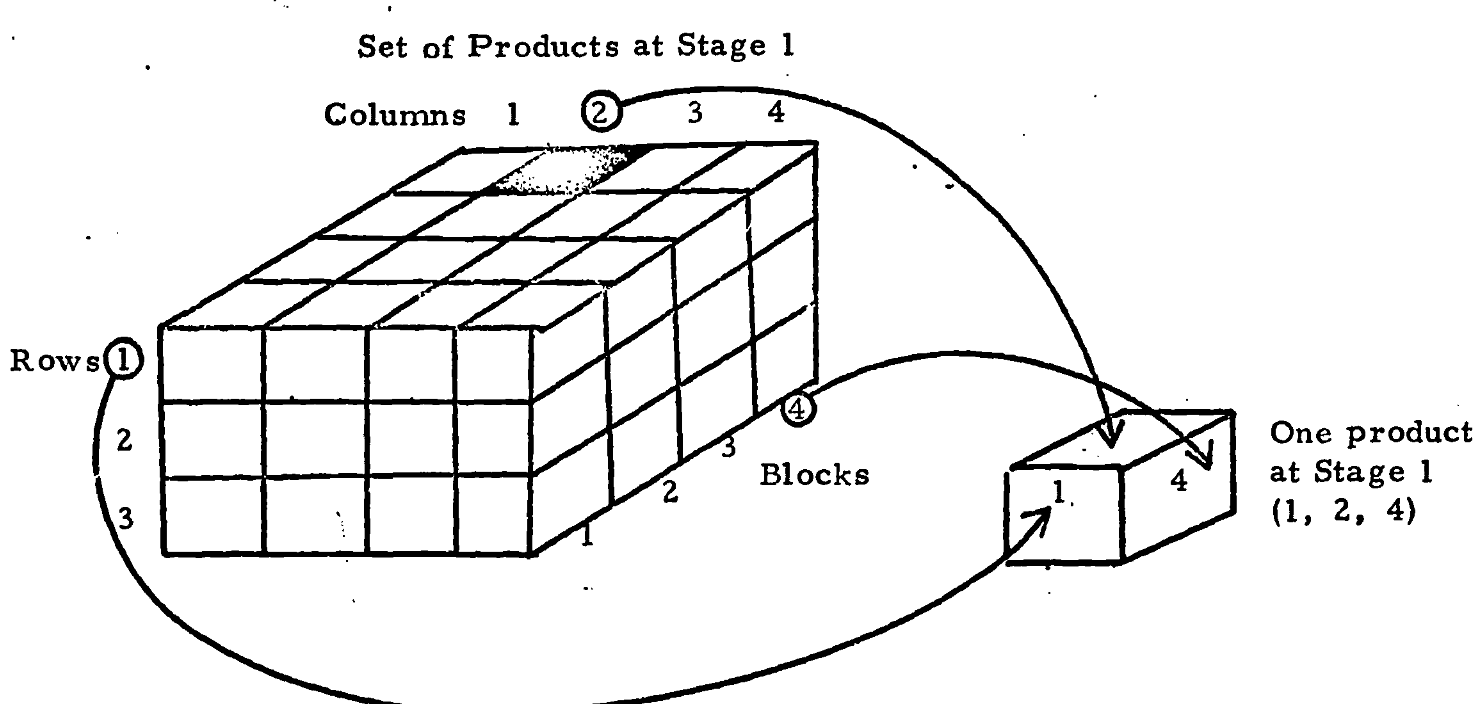
If A and B were conceived to be the two forms of a given set of products across any period of time or across a temporally defined "stage" of development, then the procedures for mapping the elements of A onto the elements of B would be those involved in the actual process of development corresponding to that stage. Hence, the specification of any of the elements of A would provide a characterization or minimal definition to the set of products at stage 1 or time 1, and the specification of any of the elements of B would minimally define the set of products at stage 2, or time 2. The flow of the set of products from its A form to its B form then could be characterized by a means of (1) a time line as illustrated below:



or (2) an R&D stage line as illustrated below:

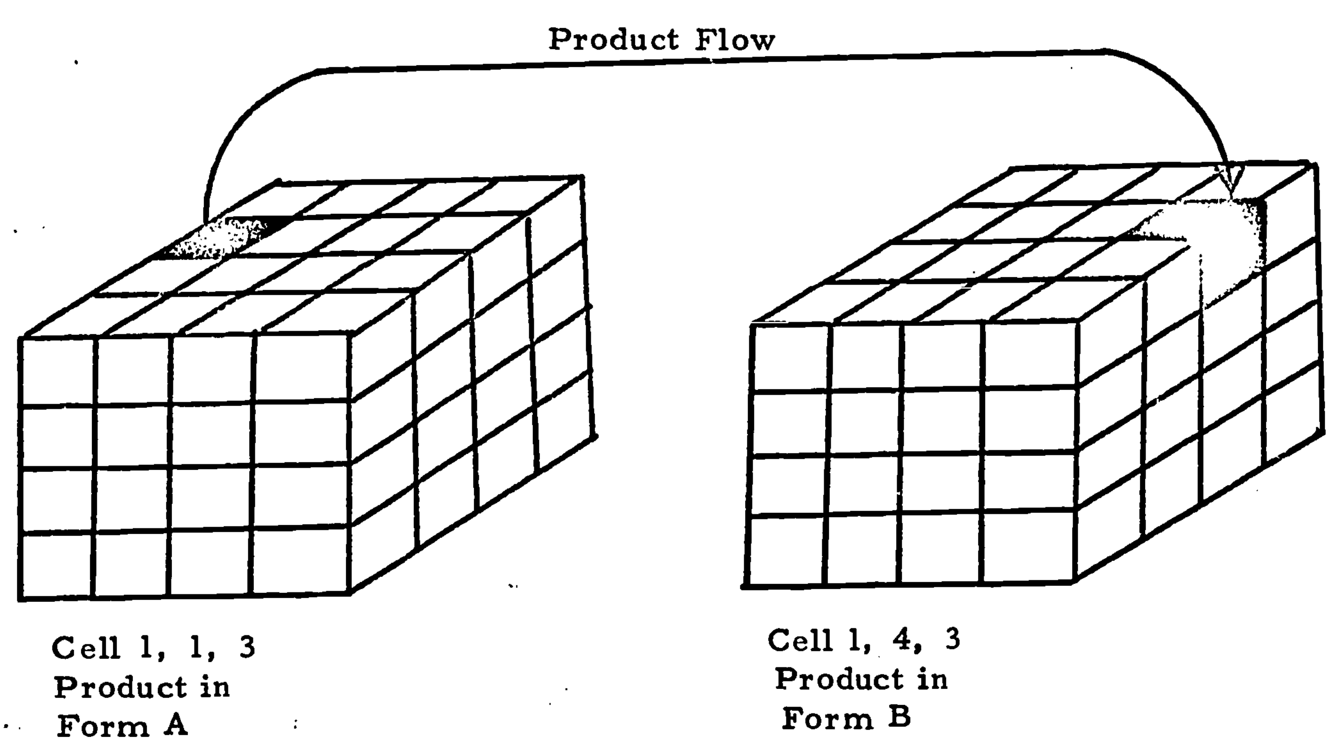


In order to illustrate the hypothetical flow of a set of products through the educational R&D process, we will attempt to show the R&D process as a mapping of the elements of the rows, columns, or blocks of one three-dimensional minimal definition matrix onto the rows, columns, or blocks of another. Thus, in the presentation any one cell of a matrix would correspond to a hypothetical product in the set, and the identification of any particular product of interest would be defined by the appropriate row number, column number, and block number for the given matrix. This is illustrated below.



The precise definition to the given product will be provided by the identifying elements in row 1, column 2, and block 4.

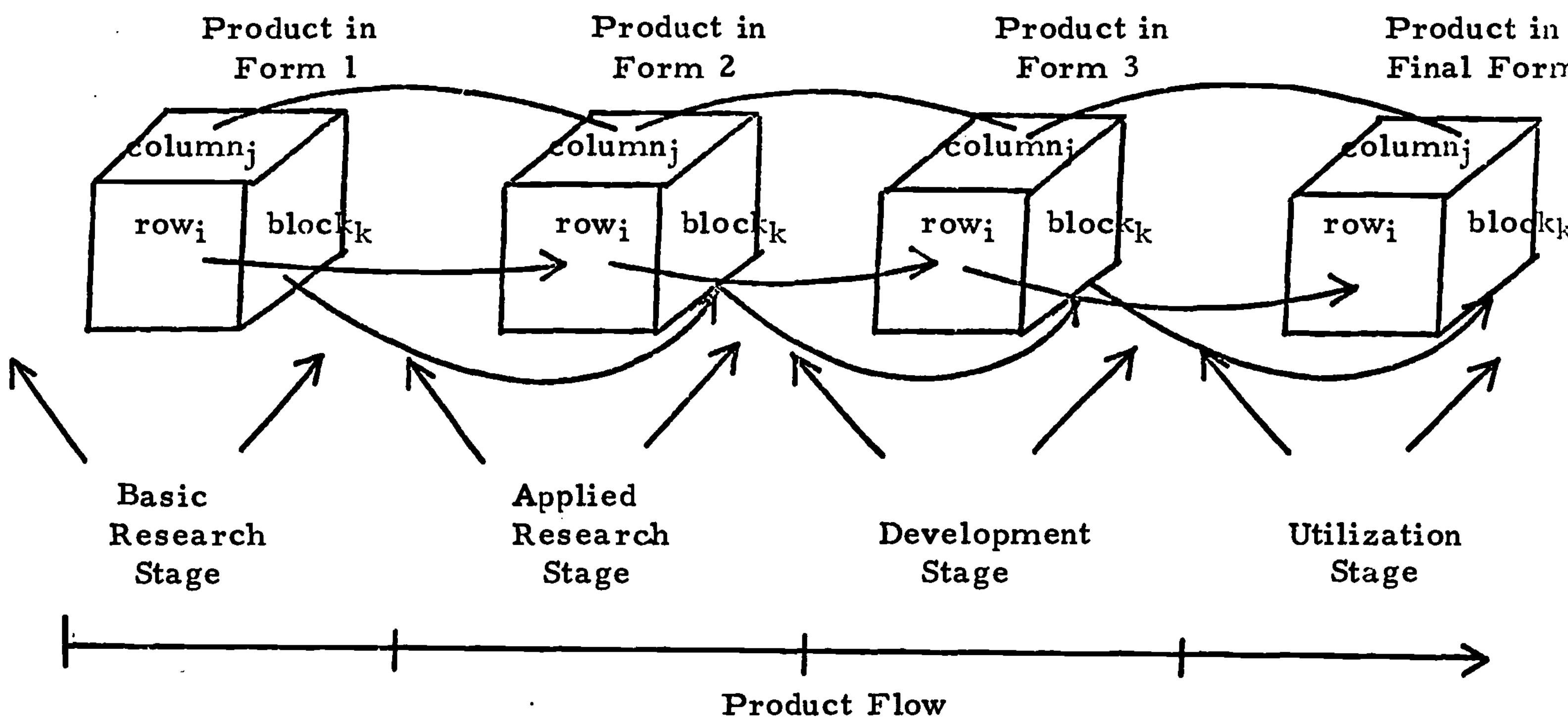
The flow of any given product through a particular stage of the R&D process would be seen as the mapping of one or more of the elements of each of the rows, columns, or blocks which define it at the beginning of a stage onto one or more of each of the rows, columns, or blocks of another which define it at the end of a later stage. In other words, the "flow" of a product is from the precise cell of one matrix onto the precise cell of another. For example, the product might be located in row 1, column 1, and block 3 at stage A, and located in row 1, column 4, and block 3 at stage B. This might be illustrated in the following way.



In summary, we will use a particular minimal definition matrix to correspond to the set of all products under consideration at a given stage. We will use a cell of the same matrix to correspond to one particular product of the set under consideration at the same stage. Two matrices would be needed to show the set of products at two different stages and one cell from each matrix would be needed to show a given product at two different stages. The R&D process for one stage for all of the products in the set could be conceived as a mapping of the matrix A at stage A onto the matrix B at stage B, and the R&D process for one product could be conceived as a mapping of one cell of matrix A at stage A onto one cell of matrix B at stage B.

If the flow of products was to be depicted for the four major stages of the R&D continuum as discussed in Part I of the paper, then four minimal definition matrices corresponding to the form of product during or after the basic research stage, during or after the applied research stage, during and after the development stage, and during and after the utilization stages would be required to show all phases of the R&D process, and the flow of any one product could be denoted by the appropriate mapping from the cell of a matrix at an earlier stage onto the cell of any later stages.

This might be displayed as shown below:



It should be noticed that the flow of products as depicted above may be either time bound, dollar bound, or stage bound. However, it may also be bound perhaps more appropriately by the specification for its form at the several stages. In other words, the time schedule for completion of an R&D stage should be more dependent on the requirements of the minimal definition matrix rather than some arbitrary time or resource allocation.

Elements and Definitions

In the presentation of the model that is to follow, the rows, columns, and blocks of these "product form" or minimal definition matrices and their respective elements will be assigned descriptive labels which are intended to correspond to the kinds of information that would need to be specified to provide a basic definition to the products at a given stage. At least a major part of the evaluator's responsibility is to help provide information relevant to the more complete specification of these elements by the development group.

Before proceeding with the explication of the model, it should be understood by the reader that although these labels, concepts, or terms have been used to attempt to define the elements of a row, column, or block of a product matrix more specifically, they are not necessarily exhaustive, mutually exclusive, or precise, rather we have sought to at least bound the general construct by the use of more specific examples, or terminology. In other words, we did not choose to refine our terms precisely for this presentation, and have assumed that four stages would be sufficient to at least span the R&D continuum, so that the reader would focus more on the general features of the model rather than its precise operational form.

Research and Development in the Physical Sciences

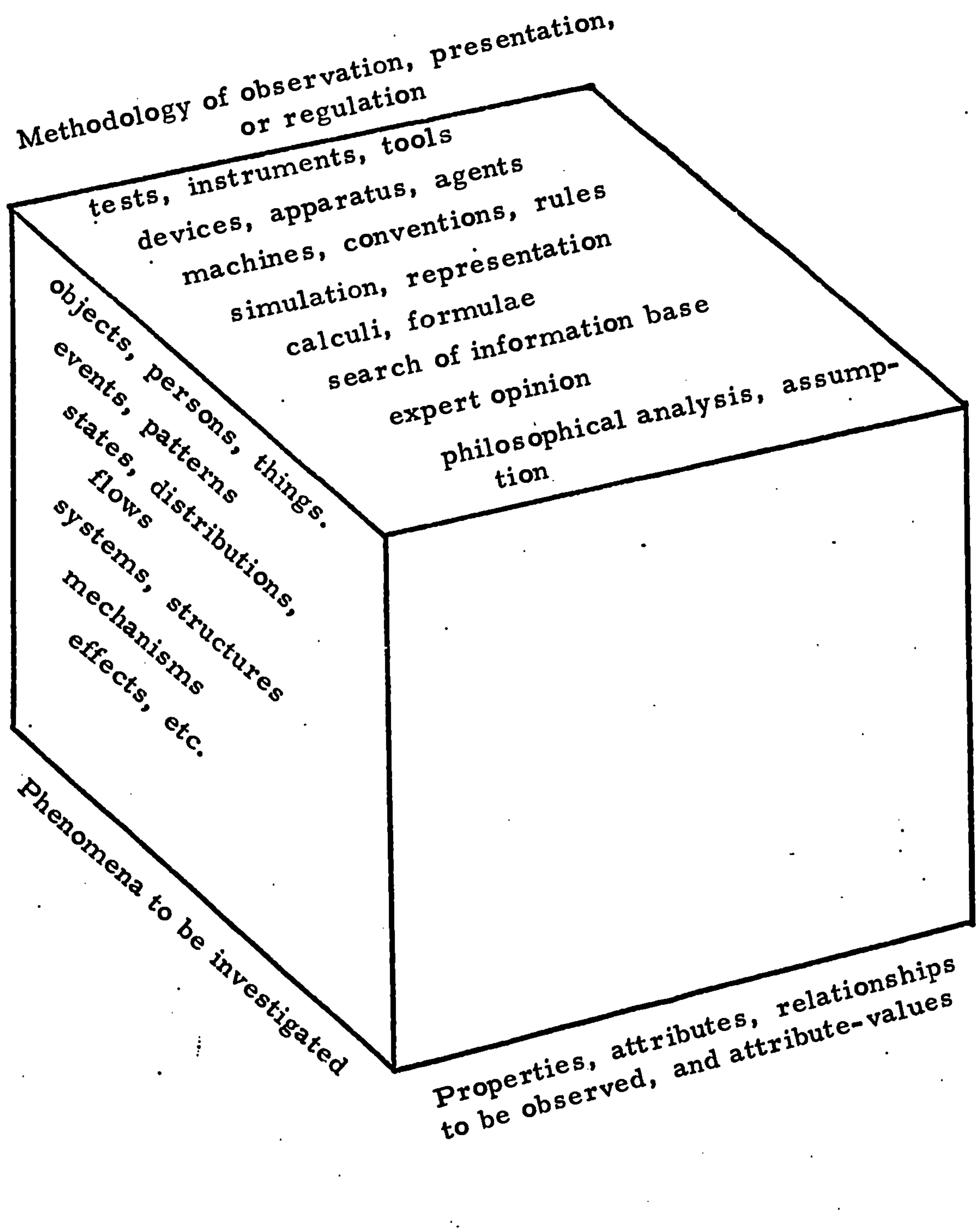
The long history of industrial research and development provides the initial set of data for the exemplification of the model. Since most industrial R&D is based upon data and knowledge derived from the physical sciences, we will begin the presentation by a discussion of the R&D continuum as it applies to this area of scientific endeavor.

We can conceive of basic research in the physical sciences as providing a research data base containing at least a minimal amount of information to be useful. This would include (1) specification of the phenomena under investigation, (2) specification of the methodology used in the regulation, presentation, exemplification, data generation, or description of the phenomena, and (3) the specification of the relations, properties, attributes, or attribute-values controlled or observed. This is illustrated in Figure 9. The purpose of this basic research in the physical sciences is seen as providing (a) phenomenal description and chara-

Insert Figure 9 about here

terization, (b) data for theory generation, fit, or revision, and (c) possible product potential. The phenomena as shown in the rows or the "i" side of Figure 9 may be objects, persons, things, events, states, patterns, distributions, flows, systems, structures, mechanisms, effects, and the like. The methodology of regulation, presentation, and exemplifications are shown in 9_j. These would include: tests, instrumentation, devices, apparatus, tools, agents, machines, conventions, rules, calculi, formulae, simulations, representations, naturalistic observation, search through relevant literature, research findings, or other information bases, expert opinion, philosophical analysis and assumption.

Figure 9. Product characterization at the basic research stages--physical sciences



The properties, attributes and attribute-values and relations to be controlled or observed (9_k) are the class of forms that the characterization of the phenomena may take.* Any given product at the basic research stage can thus be minimally defined by providing specification of these three elements for the given product.

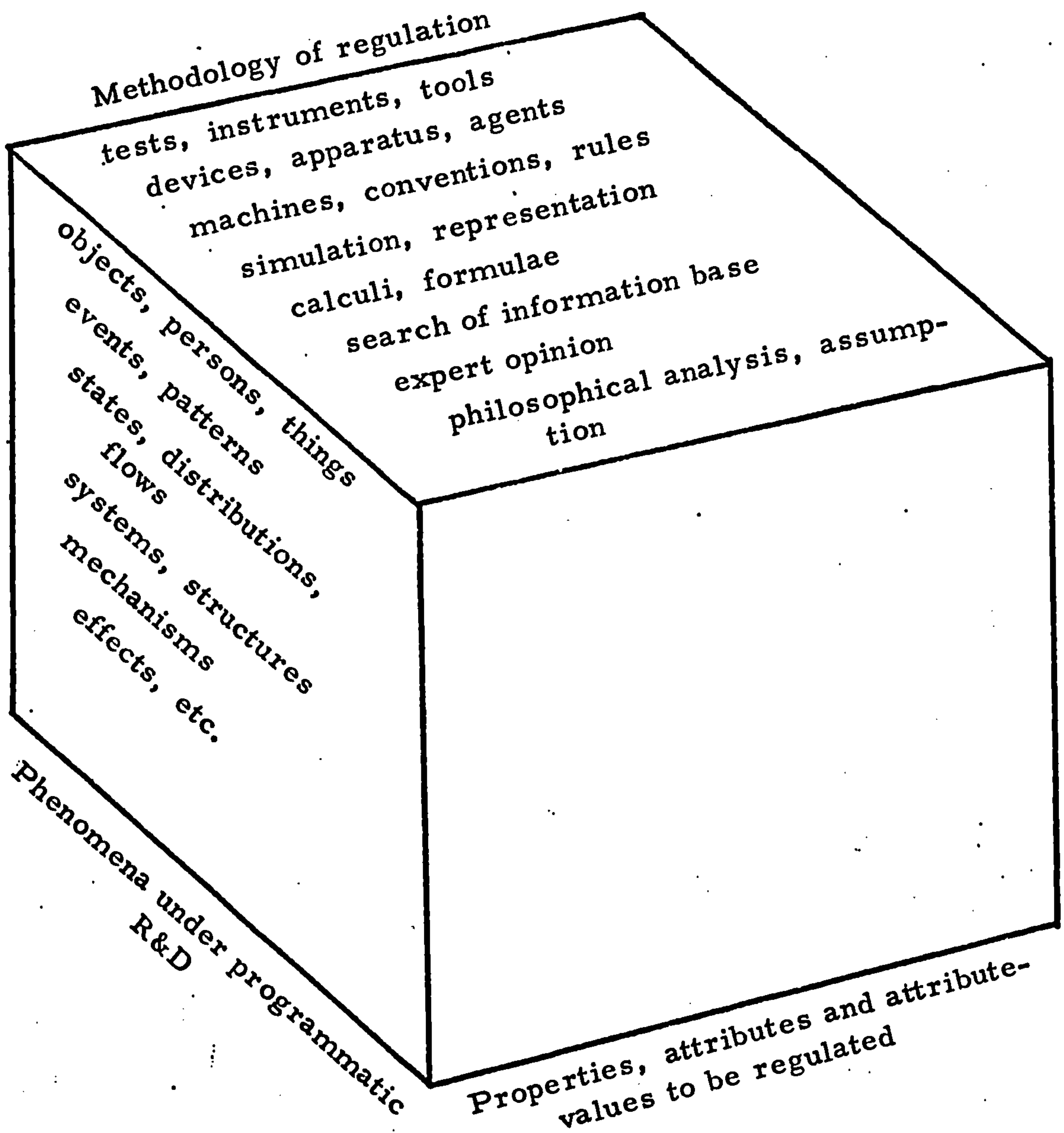
With this as a starting point, we will try to develop an idealized R&D continuum of product flow from basic research stages to the final product utilization stage that might be applicable to the physical sciences. This will require Figures 9, 10, 11, and 12 for depiction. Figure 9 has already

Insert Figure 10 about here

been described as the basic research stage. Figure 10 shows the matrix corresponding to the product form in the applied research stage. Specification of the phenomena, the methodology of regulation, and the properties are required in the definition or characterization of the product at this stage, and may be performed by a one-to-one mapping of the corresponding elements of the rows, columns, and blocks used in the basic research stage. The purpose of applied research is what distinguishes it from the prior

*The elements of the matrix are a modification and extension of those prepared by Rita Lerner of the American Institute of Physics for use in classifying research information in physics.

Figure 10. Product characterization at the applied research stage--physical sciences



basic research stage. The purpose of applied research is seen as providing (a) verification of phenomenal characterization and description, (b) phenomenal replication, and (c) explication of critical factors in methodology for subsequent illustration, presentation, and predictable regulation of the phenomenon.

Figure 11 shows the matrix corresponding to the product in the development or engineering of technology stage. At this stage the phenomena

Insert Figure 11 about here

will need to be characterized by some hypothetical user types and by the properties under present use. Specification of the methodology of packaging can be achieved by a mapping of the elements of methodology of regulation from the applied research stage 10_j onto the packaging alternatives in the development stage 11_j, and the properties to be illustrated or to be regulated by the new methodology 11_k are those which are desired for display or reliable use mapped from 10_k.

Figure 12 shows the matrix corresponding to the utilization, marketing, or diffusion stage of the R&D continuum. At this stage the product

Insert Figure 12 about here

Figure 11. . Product characterization at the development stage--physical sciences

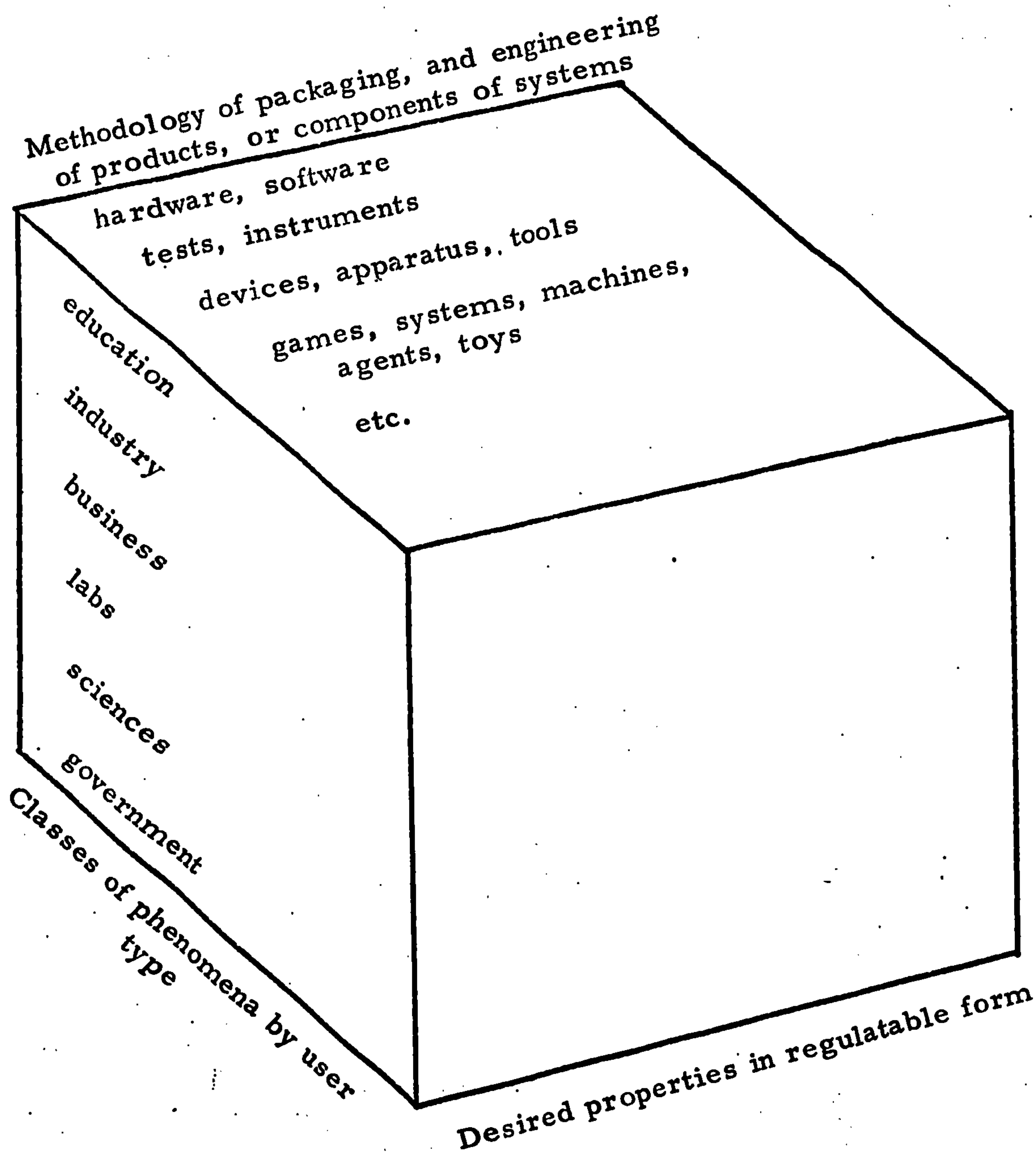
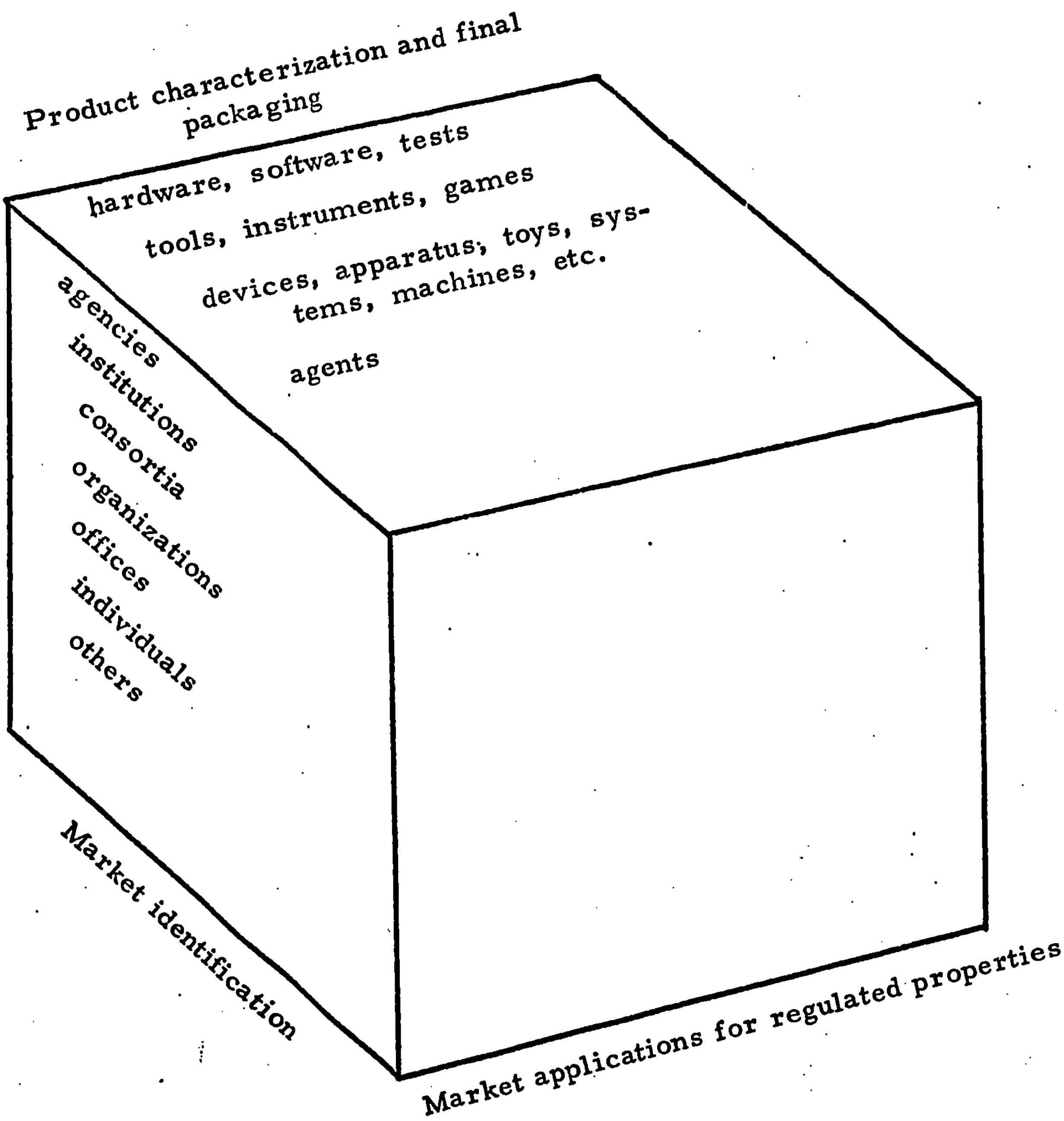


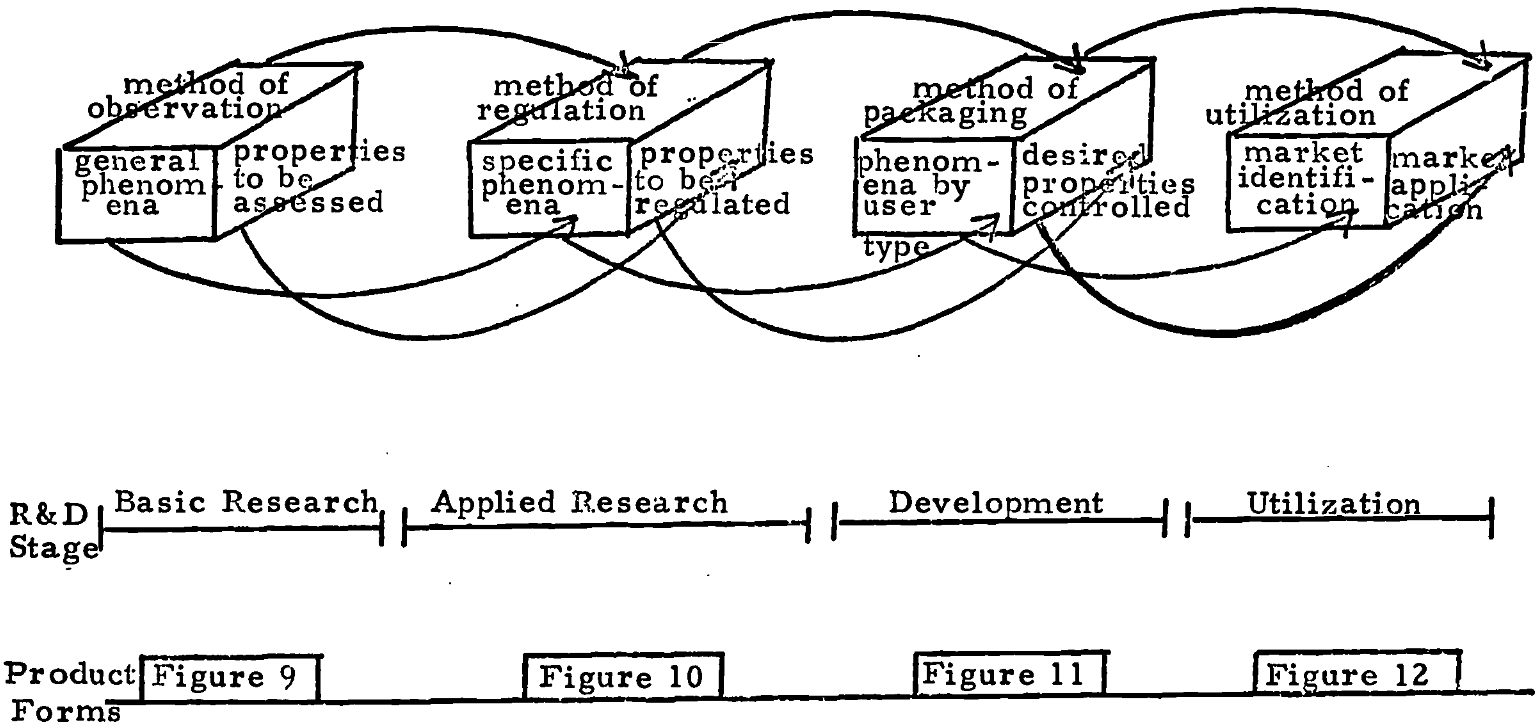
Figure 12. Product characterization at the utilization stage--physical sciences



can be defined by specification of the market to be identified, specification of the product description and final packaging, and specification of the market applications. The classification of user type from 11_i of the prior stage has been mapped onto the elements of market identification in 12_i. In a similar manner the elements of the methodology of packaging 11_j has been mapped onto the elements of product description and packaging 12_j, and the elements corresponding to the desired properties have been mapped onto the elements of market application 12_k. The basic purpose of this stage is the utilization of technology, and when this stage is completed, the R&D process terminates unless new inputs are available from earlier stages of research and development.

The major product that emerges at the termination of this process is seen as the "hard" product, the hardware, software, systems, components, machines, tools, instruments, games, devices, toys, etc. Along the line of development, however, additional products or by-products may also emerge in the form of research reports, technical documents, new or improved processes for manufacturing, etc. These "knowledge" products, process products, and types of improvements of them, are valuable in their own right, and we will come back to a more careful examination of their characteristics in a later section.

In the explication of the idealized R&D continuum for the physical sciences, we have sought to show the several most important stages of the process, and to illustrate how the elements of the product definition at one stage can be mapped onto those at a later stage to provide the basis for continuity of product form as it moves through these stages. This is illustrated below:



As shown in the preceding illustration, four discrete major product forms are required to be completed to cover the complete range of the R&D continuum. The first major product form (and the basic research stage) is completed only when there are one or more scientific products which contain the specific information required for completing the appropriate elements of the rows, columns, and blocks of Figure 10. The third major product form (and the development stage) is completed only when there is the first "hard" product available, and when there is documentation or scientific products containing the specific information for completing the appropriate elements of the rows, columns, and blocks of Figure 11.

Obviously if this development stage has actually been entered after the activities at the two previous stages have been completed, then there will be a trail of other documentation products tying the hard product to theoretical, and/or empirical data obtained at these earlier stages. The fourth major product form (and the utilization stage) is completed only when there is some documentation in product (e.g., a market analysis) available which contains the specific information required to complete the appropriate elements of the rows, columns, and blocks of Figure 12.

The idealized R&D continuum for the physical sciences is seldom accurately followed in practice. Work may begin at any stage of the continuum. There are frequent loops back between the various stages. The continuum may be truncated, and many abortive product forms occur.

In the next section we will try to develop an idealized R&D continuum of product flow from the basic research stage to the final product utilization stage that might be applicable to the social and behavioral sciences. This will require some modification of the description of the elements required for the physical sciences.

R&D in the Social and Behavioral Sciences

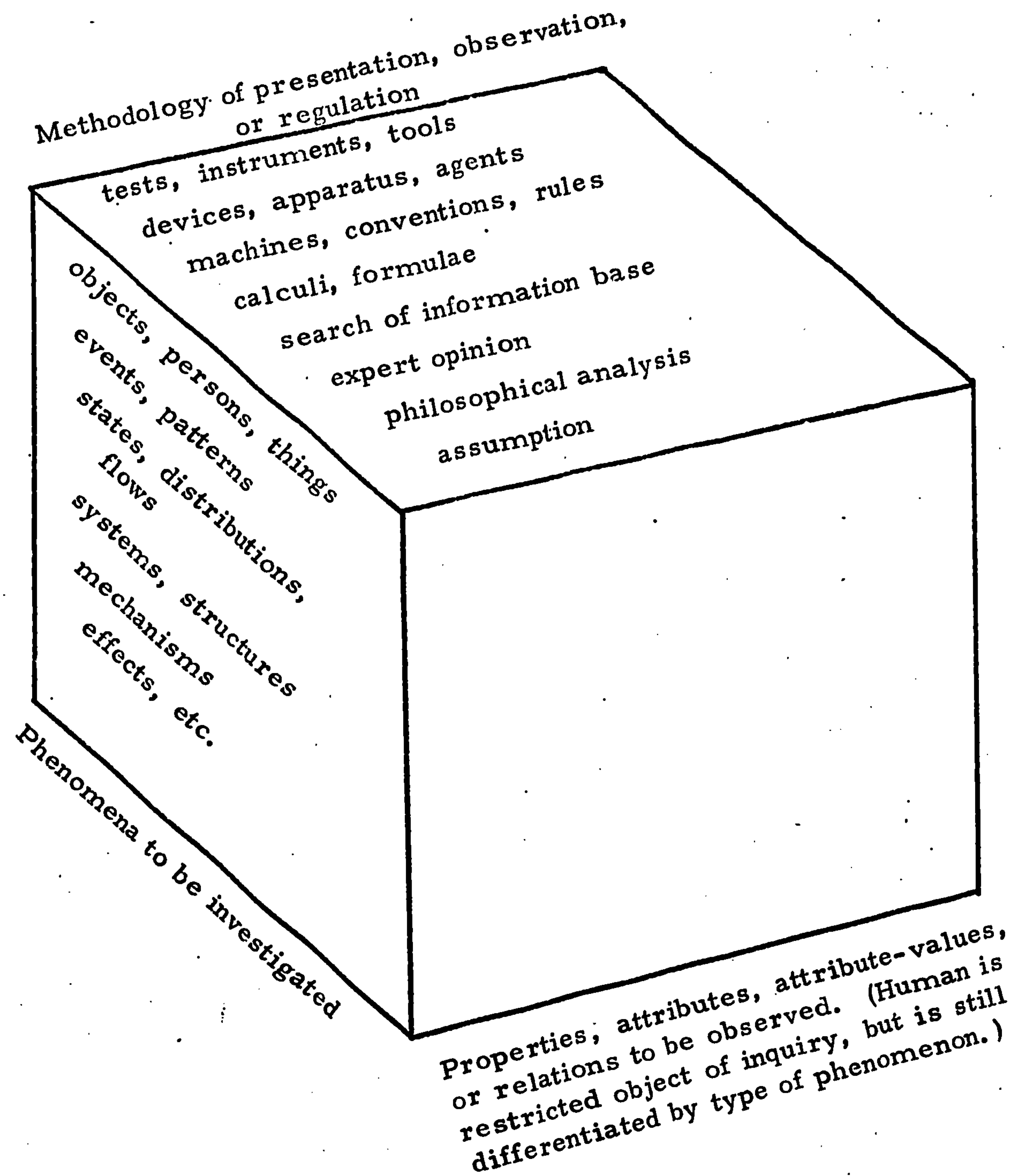
The R&D continuum for the social and behavioral sciences is shown by means of a series of matrices as shown in Figures 13, 14, 15, and 16. The basic research stage in the social and behavioral science is shown in Figure 13 and it shows the product definition as being derived from a mapping from the elements of the rows, columns, and blocks of the correspond-

Insert Figure 13 about here

ing matrix for the physical sciences. This lateral mapping has been essentially one-to-one in terms of the types of forms of the phenomena, however, it should be understood that there is now considerably greater reduction or restriction on the kinds of phenomena that might reasonably fall within the boundaries of the social and behavioral science disciplines as legitimate inquiry, i.e., we have moved from a more general level of abstration to a more specific level. In effect, we have truncated or divided the set of disciplines to which, theoretically, at least, all of the elements of 9_i could be mapped and have gone to a very restricted subset of the universe of phenomena, those which are human, for a more specific application. More accurately, the elements of $9_{i,k}$ have been mapped onto a subset of themselves.

The elements of the methodology of presentation, regulation, and illustration are derived from a one-to-one mapping from 9_j, as are the

Figure 13. Product characterization at the basic research stage--social and behavioral sciences



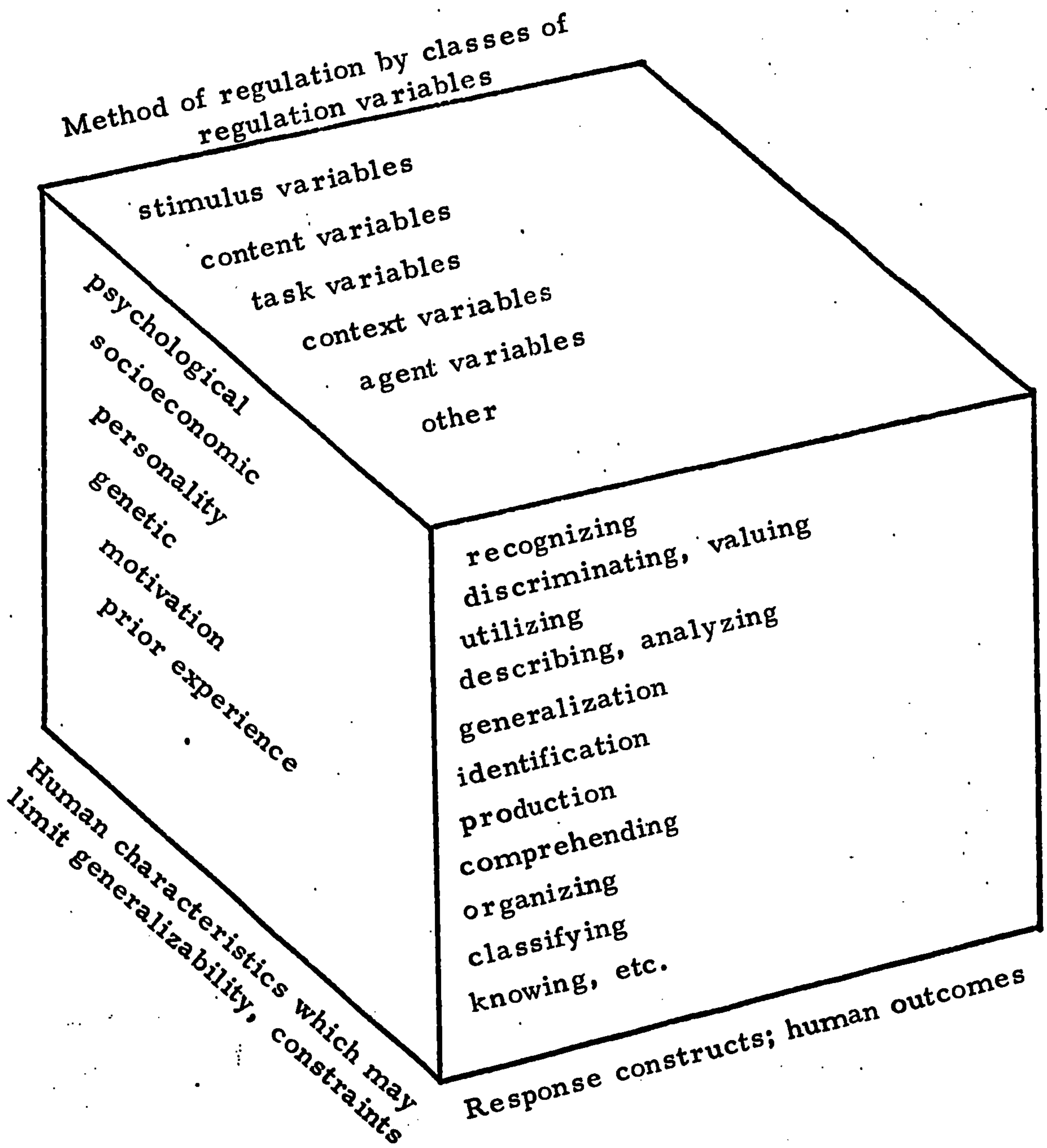
properties, relations, attributes, and attribute-values from 9_k for the truncated set. Thus, it can be seen that the forms of the products at the basic research stage for the social and behavioral sciences are quite similar to that in the physical sciences, and require the same types of information to be specified to provide a minimal definition or characterization.

In the applied research stage of the R&D process for the social and behavioral sciences, we have tried to depict the elements of the product definition matrix in the idiom of these fields in Figure 14. Because the

Insert Figure 14 about here

phenomena relevant to this discussion are essentially and exclusively human, human related, or human-like, a new set of elements will be devised by performing a classification of these phenomena according to the relevant other characteristics of humans. This set is a substitute for the elements of the set corresponding to the forms of the phenomenon in the basic research stage 13_i and denotes some possible constraints for the developer to take into account. The elements of methodology of presentation and regulation from 13_j have been mapped onto the elements of the set of regulating variables depicted in 14_j, with the elements in 14_j reflecting the explication of a hypothetical set of critical factors in the regulation of educationally-relevant behaviors. The elements which make up the regulating

Figure 14. . Product characterization at the applied research stage--social and behavioral sciences



variables at the applied stage represent the joint set of the elements of 13_i and 13_j at the basic research stage as these tend to be grouped together more generally by researchers in the social and behavioral sciences.

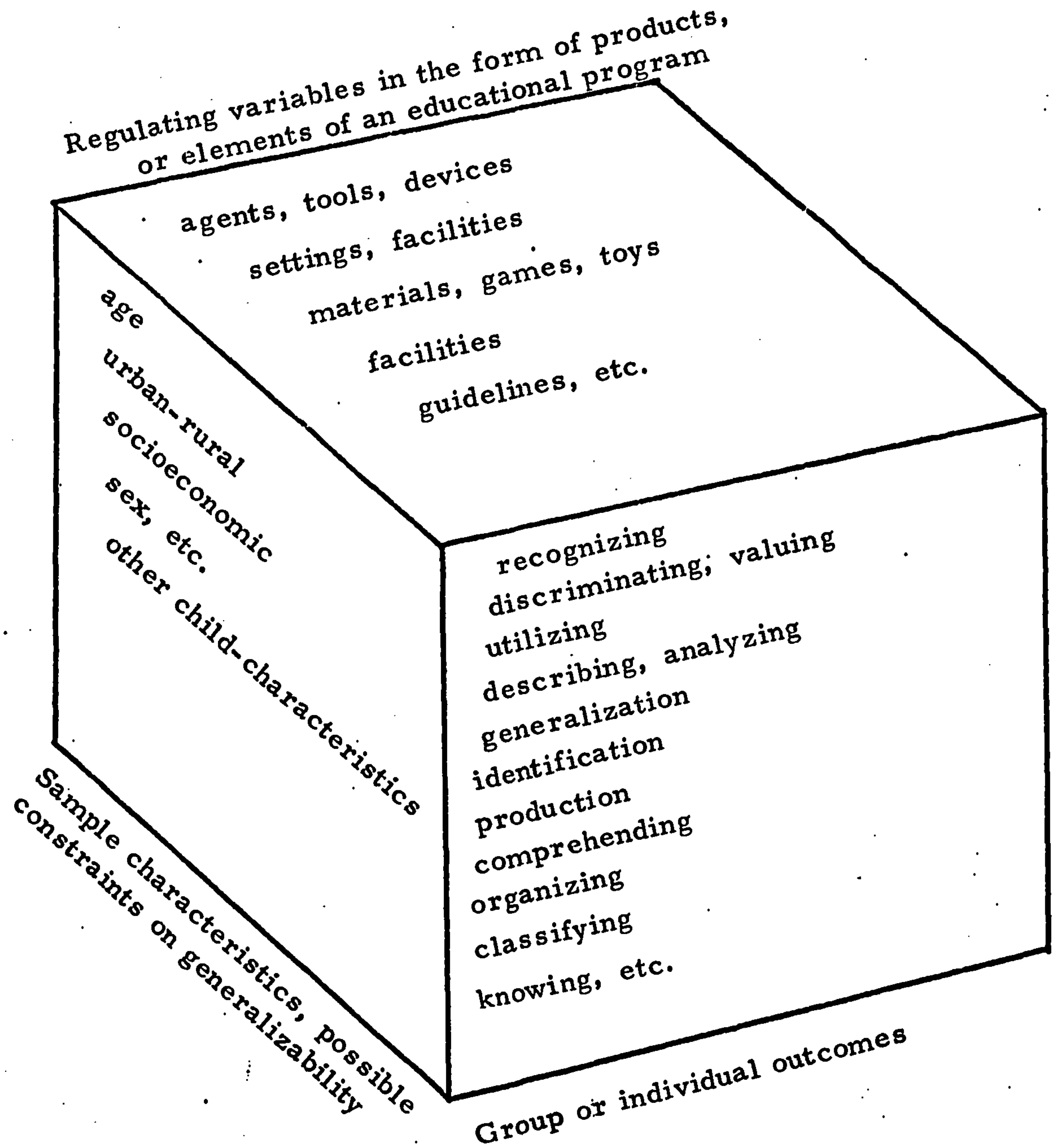
The blocks of the applied research matrix for the social and behavioral sciences represent the elements corresponding to the subset of response constructs or human outcomes of special interest to educational developers. These are seen as being directly mapped from 13_k onto 14_k. Again, we should point out that the labels used for these elements reflect an attempt to bound the general construct such as "regulating variables," rather than to give these general terms precise taxonomic definition. However minimal definition of an R&D product will require detailed information about elements on all three sides of the matrix, thus some further specification is required for a given application. The purpose of applied research in the social and behavioral sciences is substantially the same as in the physical sciences, i. e., to replicate, verify, or regulate the phenomenon in terms of its salient properties.

At the development stage of the R&D process in the social and behavioral sciences, shown in Figure 15, the product takes on a form more

Insert Figure 15 about here

familiar to those of us in educational laboratories and R&D centers. At

Figure 15. Product characterization at development stage--social and behavioral sciences



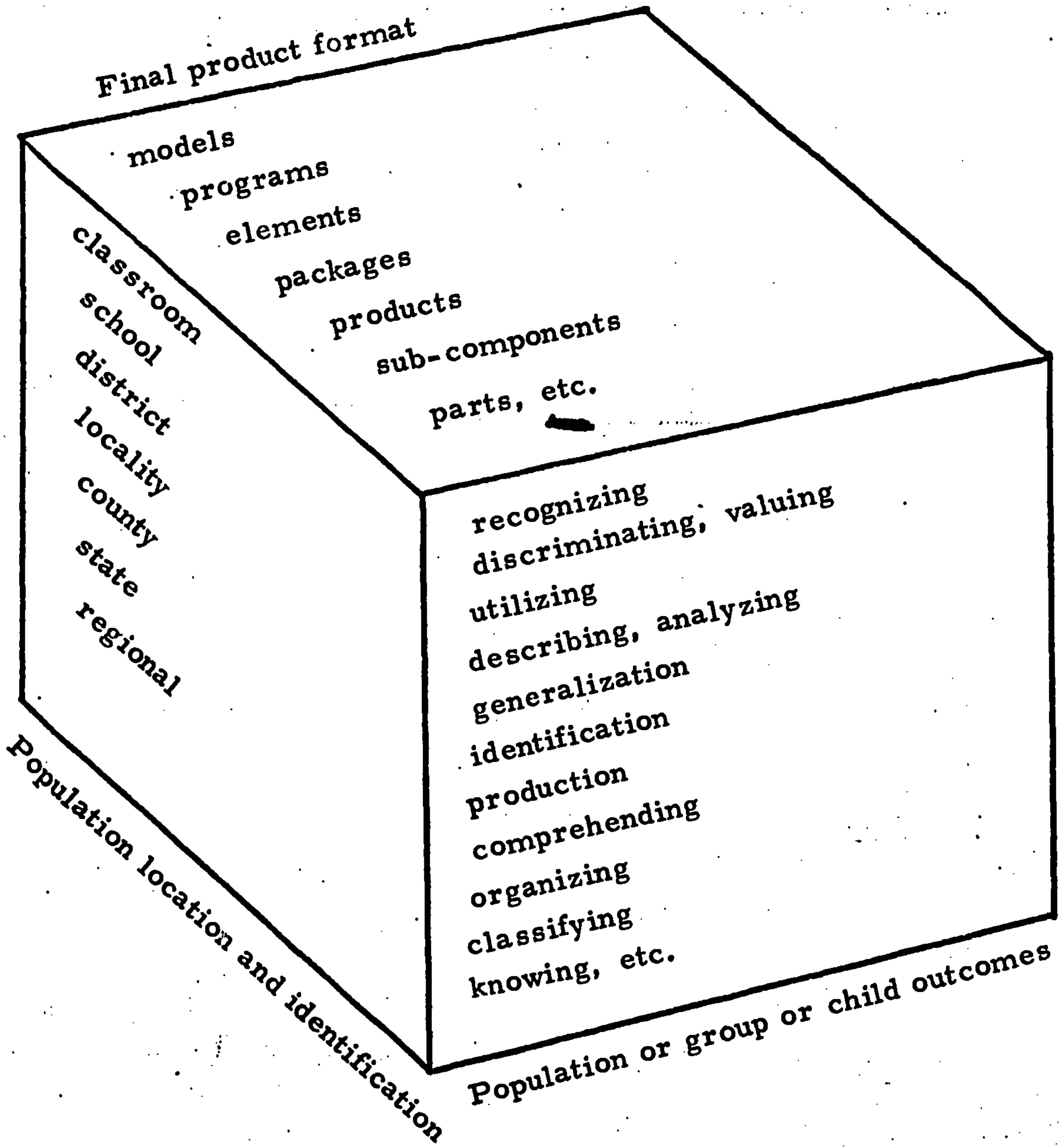
this stage the product requires specification of the characteristics of the sample of potential users or consumers to whom it is potentially applicable which can be mapped from 14_i; the method of packaging or the means for the conveyance of the regulating variables which can be mapped from the regulating variables in 14_j, and reflect the forms of 11_j as a lateral mapping would show; and the group, child, or other outcomes desired are on the blocks, and these have been mapped from 14_k. At this stage it is clear that the forms of the product in the social and behavioral sciences can be depicted as "capturing" or "containing" the regulating variables explicated earlier in the process. Thus, any adequate product characterization would seem to require the specification of the major regulating or independent variables that are seen as becoming operational through it, specification of the group, child, or other outcomes mediated by it, and any limits on generalizability for any effective evaluation to be possible.

At the final stage of the R&D process, as shown in Figure 16, the product definition takes on information concerning the population character-

Insert Figure 16 about here

istics mapped from 14_i, the product forms mapped from 14_j into a packaging format as a program, a component of a program, or remaining as a discrete product; and specification of the population, group, or child outcomes

Figure 16. Product characterization at utilization stage--social and behavioral sciences



desired mapped from 14_k.

The products that emerge at the conclusion of the R&D process in the social and behavioral sciences are of several kinds. They would generally include the "hard" products of R&D in the physical sciences, and the knowledge products as well, but the social and behavioral sciences would also produce some special classes of products including (a) anthropoactive products, i.e., those which focus on the use of human agents for the regulation of critical variables, and (b) information products, i.e., those products which focus on the capturing, digesting, repackaging or display of data or information about the knowledge base, to a person or to a user group. Each of these latter two types of products will be discussed in the paper in some detail.

Further clarification, the R&D continuum for both the physical sciences and the social and behavioral sciences is illustrated in Figure 16a.

Insert Figure 16a about here

The flow of products in both scientific areas reveal some important similarities of elements at certain stages, and of the products. For example, the phenomena manifest themselves in comparable forms, methodologies are conceptually similar, etc. It should also be noted that the boundary line for the R&D stages are not time-determined at all. They are primarily

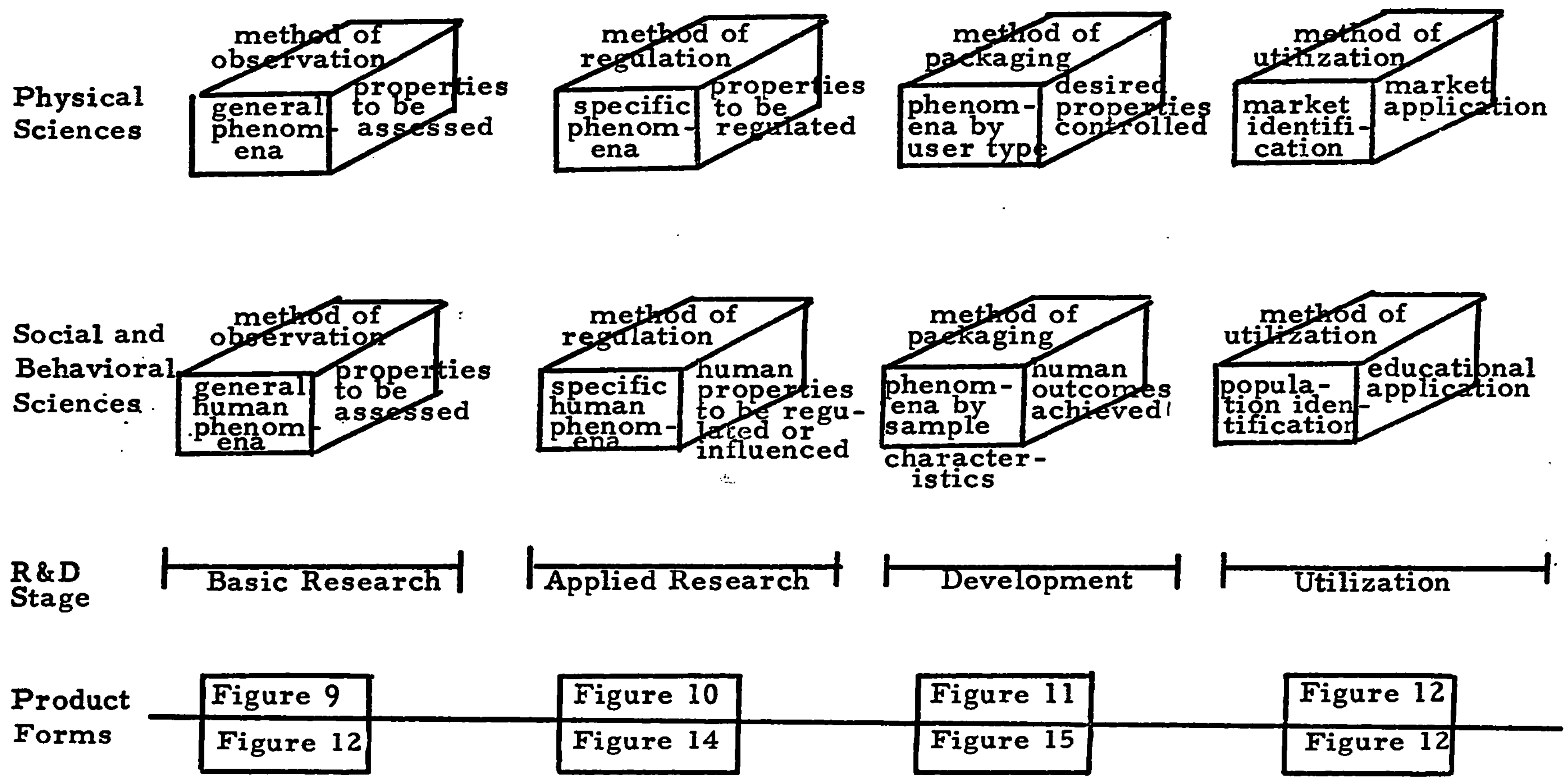


Figure 16a. Comparison of the products forms at the four stages of R&D continua for the physical and the social and behavioral sciences

determined by the availability of the information required in the minimal definition of the product at the given stage. The reasons for presenting the continua for both the physical sciences and the social and behavioral sciences side by side was to help point out some similarities in the elements contained in each, as well as some transformations of labels that help to provide a handle to the form of the product as it moves through the stages of the R&D process in both areas.

The R&D Program in a Laboratory or Center Setting

The nature of the relationship between the product qua product and the R&D process as it applies to the programmatic R&D activities of one of CEMREL's programs, the National Program on Early Childhood Education (NPECE) is illustrated in Figures 17, 18, 19, and 20.

The planning for the program begins at the Program Design and Planning Stage (which is concerned with the selection of a means for the solution of a major national problem) that of providing for early childhood education. This stage is seen as a minimal definition matrix requiring (a) specification of the child development objectives, and management objectives for early education programs, (b) specification of the elements and products needed as a means to attain these objectives, and (c) specification of the general activities that will need to be undertaken to develop the elements and products required in the program, and the organizational structure for the support of these activities which can then be derived from analyses of these functions.

Insert Figure 17 about here

The basic research stage for a program such as NPECE is seen in Figure 18 as a matrix requiring specification of a subset of R&D activities

Insert Figure 18 about here

or functions to be performed on the rows (derived from the analysis of the processes of R&D presented earlier in the first section of the paper, see Table 2), specification of the program components needed, and of the methodology of presentation and regulation obtained by a one-to-one mapping from 13_j, and block elements corresponding to the properties, relations, attributes, and attribute-values of child-related phenomena derived from a subset of 13_k.

In the applied research stage, the elements corresponding to another and subsequent subset of R&D activities can be mapped from 17_i, the regulating variables can be mapped from a subset of 14_j, and the set of

Insert Figure 19 about here

Figure 17. Scope of National Program on Early Childhood Education program design stage

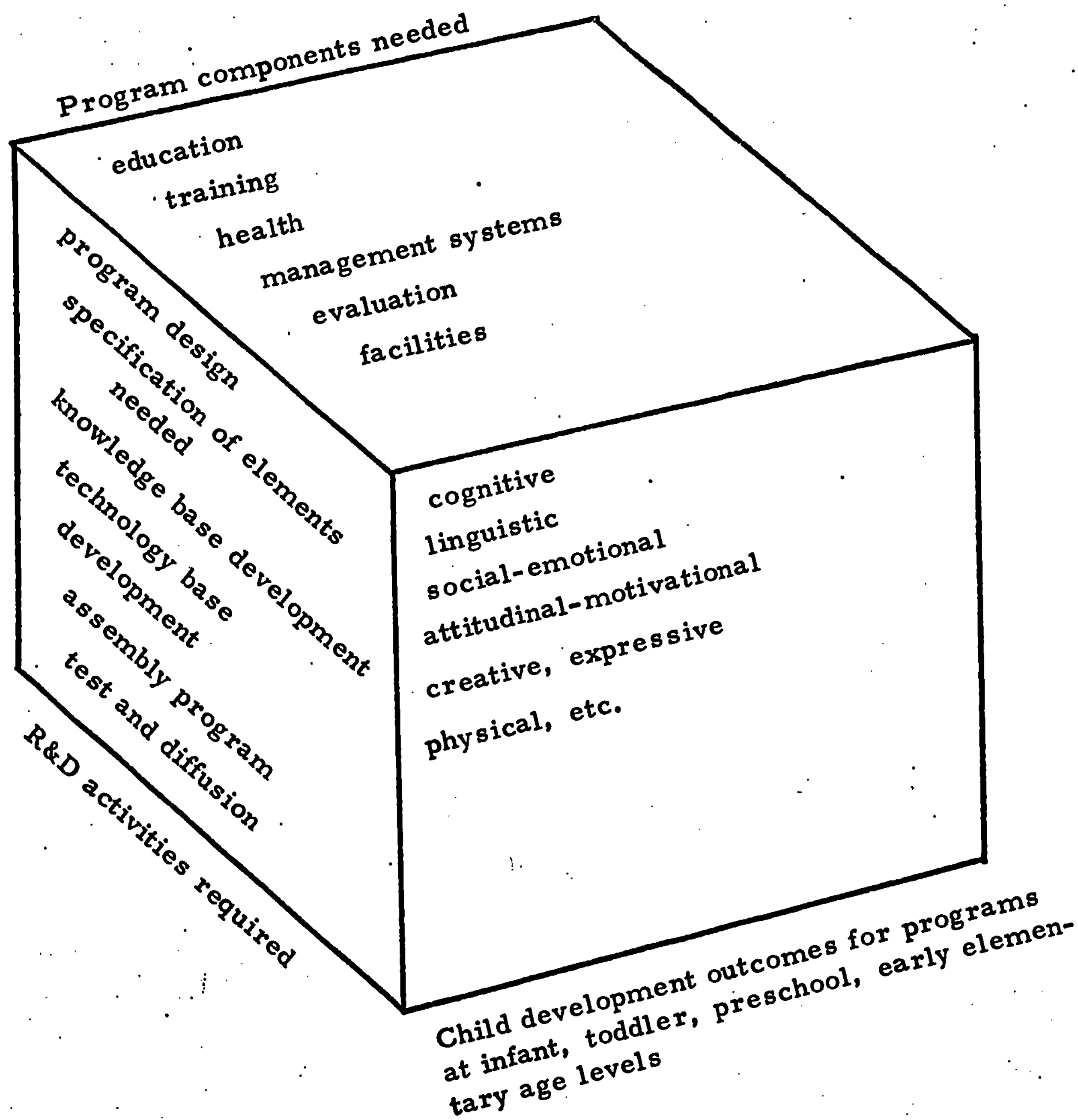


Figure 18. Program objectives--basic research stages (Focus on knowledge base development)

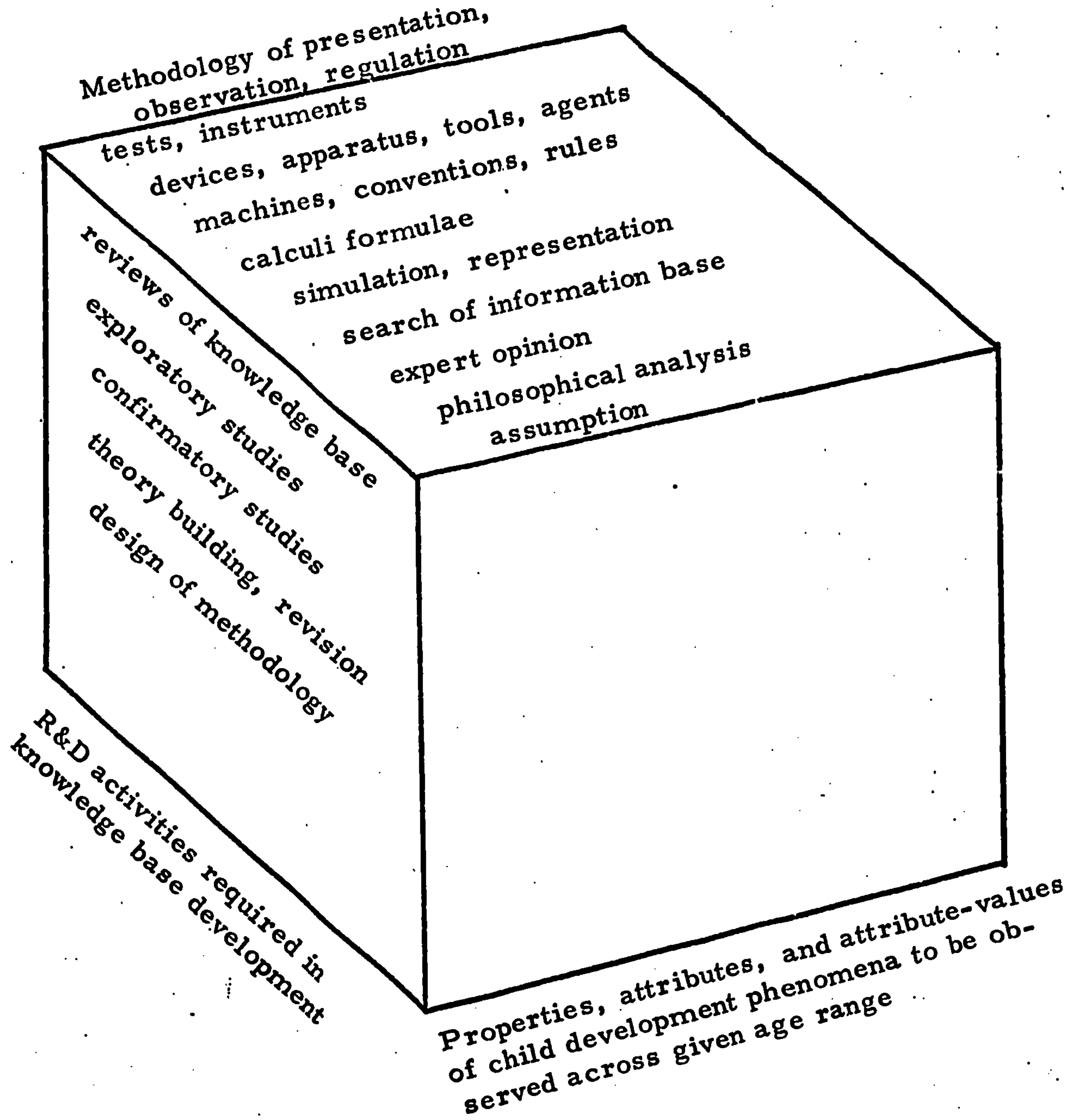
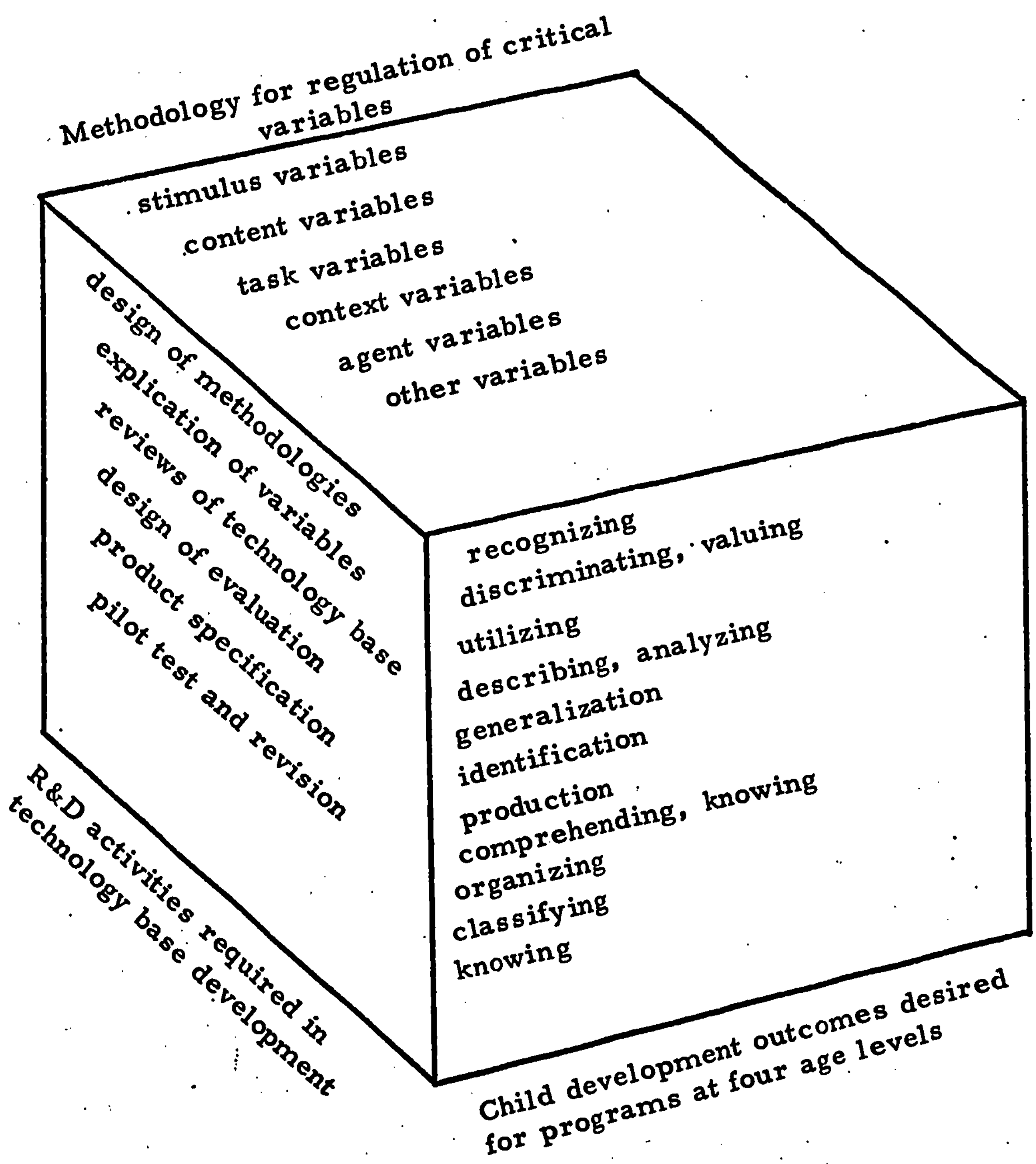


Figure 19. Program objectives--applied research stage
(Focus on development of technology base)



child outcomes are mapped from a subset of the response constructs in 14_k. In a similar manner the set of products at the development stage of the R&D process for NPECE as shown in Figure 20 include activities mapped from 17_i, types of formats for regulating variables mapped from

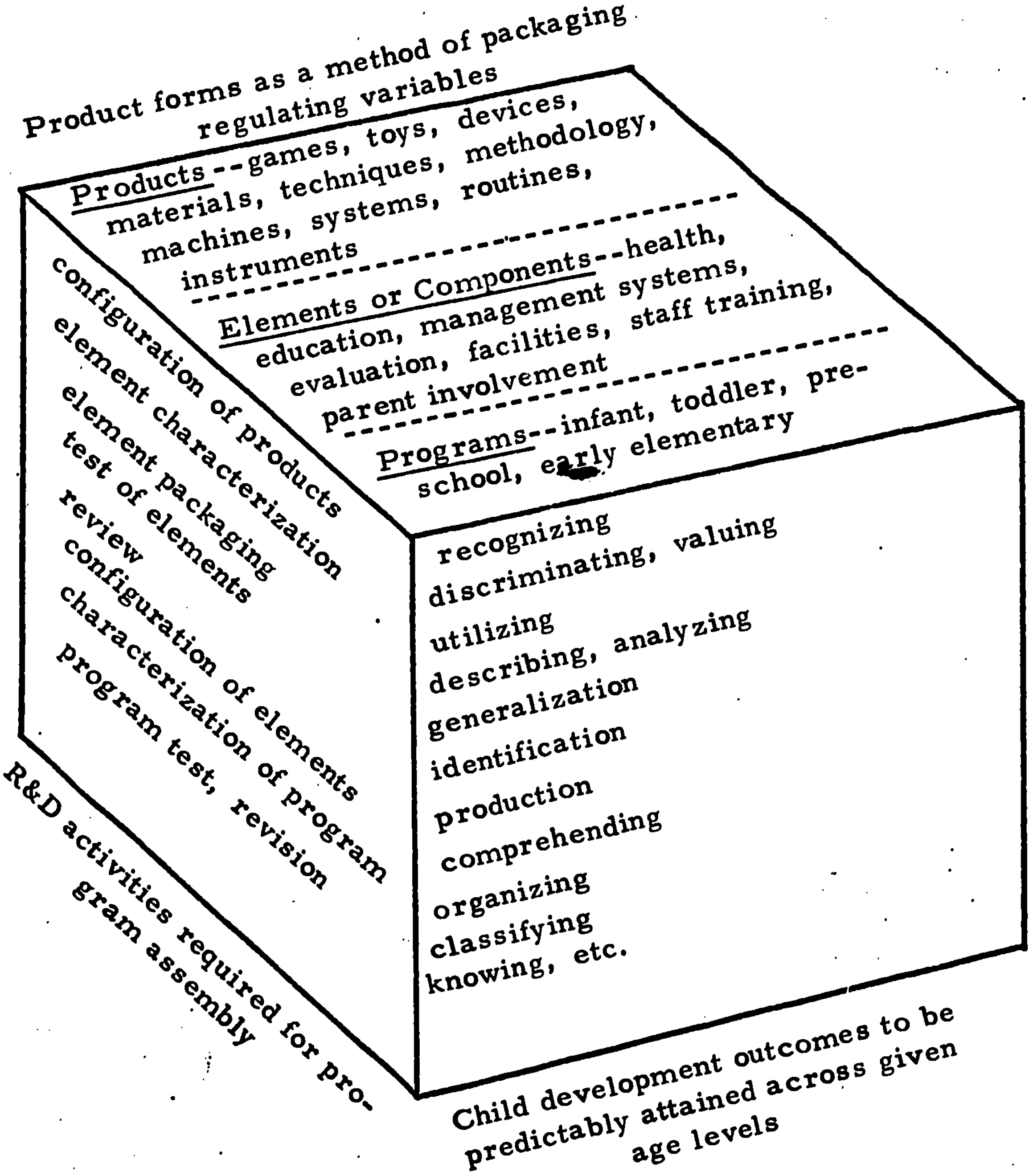
Insert Figure 20 about here

14_j, and child or group outcomes mapped from 19_k. The utilization stage would be similar to that in 15_{ijk}.

The differentiation of the required R&D activities or functions at the several stages is extremely useful as it provides a basis for deriving an organizational structure for NPECE directly under the assumption that the form of any organization should follow the specification of the functions it is created to provide.

The basic tenet which guides the work of the National Program is that educational programs today must be comprehensive in scope. They must attend to the development of many cognitive, language, social-interpersonal, and attitudinal-motivational competencies in young children. They must provide program components containing the necessary elements of care giving, instruction and curriculum, staff training, parent involvement, community support, facilities, program support and management, and program linkages which conceptually and operationally tie programs for one

Figure 20. Program objectives--development stage (Focus on program assembly)



age group to the next in order to support and maintain specific child development outcomes over a period of time; and they must also include systematic evaluation procedures in order to assess short- and long-range effects on the children they serve and as a guide to policy decisions and planning.

Program Evaluation

These goals represent the hoped-for outcomes of the National Program on Early Childhood Education over the long haul. A planning and design matrix for the range of types of early childhood programs is displayed in Figure 21. The rows contain elements corresponding to the agency of facilitation*

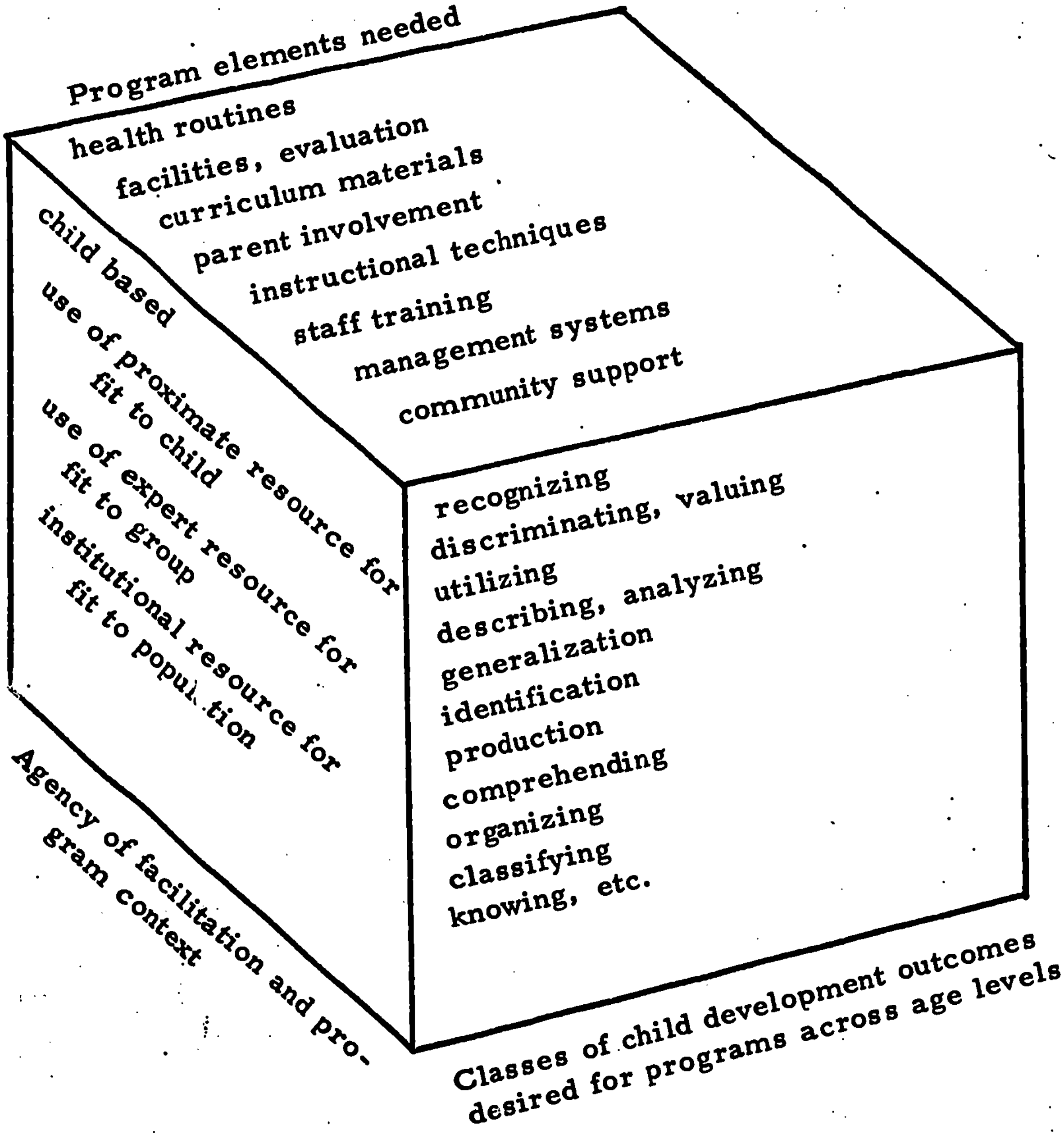
*The use of an agent of facilitation to help "fit" the regulating variables to the individual child or the particular group of children is seen as necessary for several reasons. First, the cost of development will be increased many fold, and take a much longer time period if products are required to be "consumer ready" for consumer groups of children of early age levels. Thus, fitting agents may be used to save on costs, and to speed up the development process. Second, the variability of children at these age levels on any requisite skills is likely to be normally distributed making assumptions about user "readiness" as a pre-requisite less important where fitting can be tried, or where trials can be extended or paced by the agent. Third, formal education and day care programs for young children tend to be very labor-intensive to meet state guidelines on staffing and services. This inflates the costs of providing services, and often restricts the use to those who can afford it easiest. The use of agents of different levels of skills and professional training allows scarce and more expensive talent to be used only where they are absolutely needed thus helping to reduce costs and expand the group of potential recipients of any program benefits. Other benefits might be explicated as well, particularly the opportunity for more systematic interaction of the child with one or both parents, when home-based programs are used extensively.

Insert Figure 21 about here

and the context for education and care programs of children. This would include (a) child based or child-as-agent programs; (b) programs which use a proximal resource (probably the parent or a peer) for fitting program components to a particular child; (c) programs which use an expert resource for fitting program components to a group of children, e.g., a home visitor, a visiting nurse, etc., etc., and (d) programs which use an institutional or centralized resource for fitting program components to a broader section of the population of children or to several groups, e.g., a day care center.

The program components that we have denoted on the rows of this matrix are conceived to be the classes of important elements to consider in designing any program, and these elements form a set of regulating variables which are mediated or maintained by the products which comprise them. In other words, attention to selected features of the facilities is seen as necessary because certain aspects of the facilities, e.g., <u>heating and cooling plant,</u> mediates the operation of a set of one or more important regulating variables, e.g., <u>temperature</u>, and this in turn affects <u>pupil comfort</u> directly. These elements then provide the means to effect the levels of one or more of the child development outcomes desired, specified on the blocks. Any complete program evaluation would obviously require

Figure 21. Scope of NPECE Programs



detailed specification of the regulating variables contained in all of the separable components of a program, as well as the multivariate set of child outcomes each is designed to effect. Thus, program evaluation can be conceived as a more complex version of element or product evaluation involving multivariate sets of regulating variables (one block for each element or product), and multivariate sets of outcome variables (one block for each element or product) and where each outcome is assumed, hypothesized, or previously observed to be causally related to one or more of the regulating variables of the set within a given block. The analysis of evaluation data from such highly specified programs, and the designs for obtaining the data have yet to be worked out in their entirety, but some multiple regression approach seems to be the most sensible at this time. A minimal design is schematized below:

Program Elements	Child Development Outcomes 1 2 3 4 . . . . N
a	1 0 1 0 0
b	0 1 1 0 1
c	1 1 1 0 0
d	. . . .
.	. . . .
.	. . . .
.	. . . .
N	1 0 1 1 1

1 = outcome related to element <u>a priori</u>
0 = outcome not related to element <u>a priori</u>

Anthropoactive Products

Earlier a reference was made to the development of anthropoactive products which required the use of human agents for the regulation of the critical variables. These products normally take the form of a training program or procedure for professionals or paraprofessionals in some areas and include the sets of manuals, workbooks, units, activities, courses, events, materials, etc., etc., which comprise them. A minimal definition matrix for the characterization of these products is shown in Figure 22. The elements corresponding to the variables to be regulated

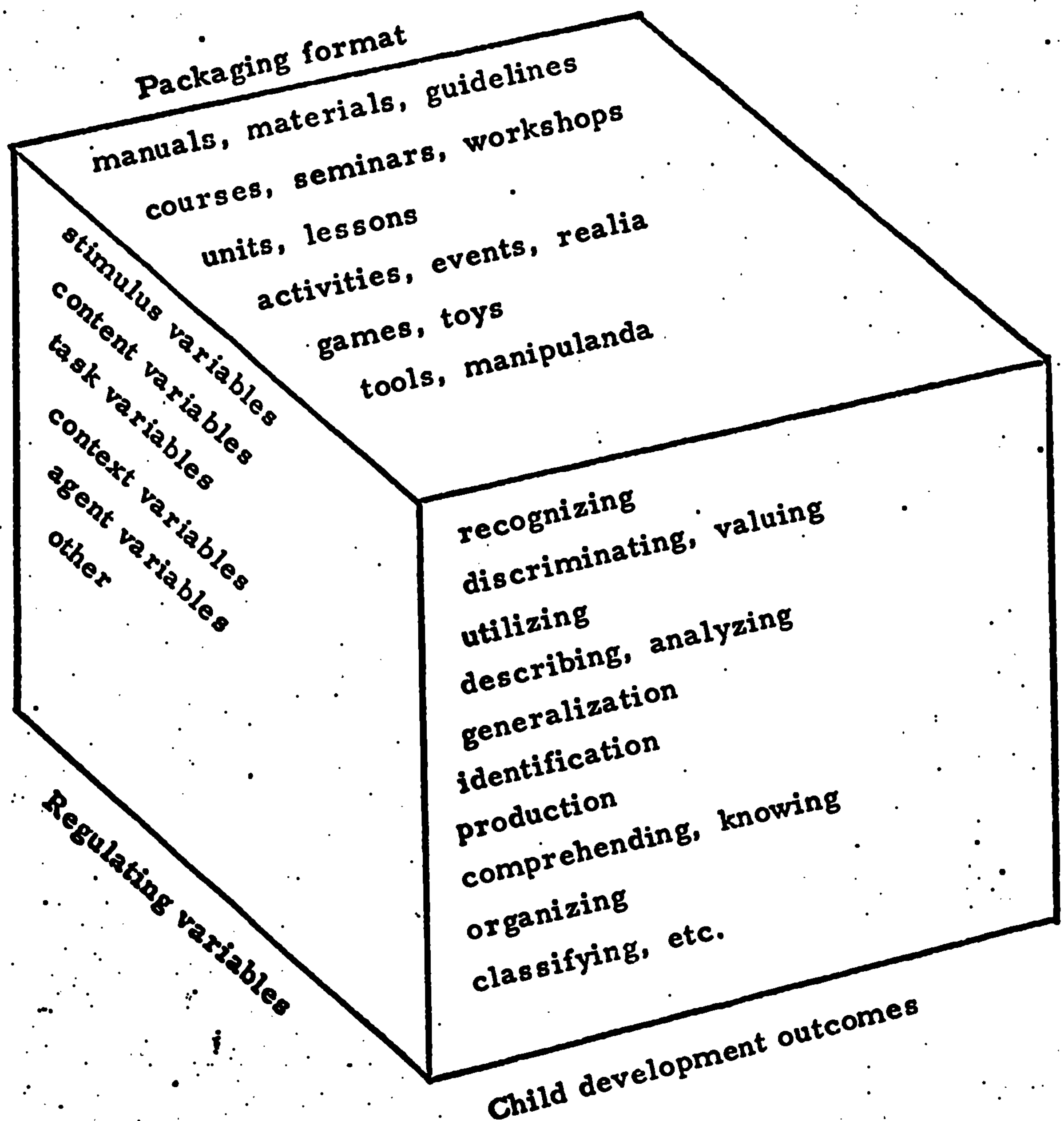
Insert Figure 22 about here

through the training program are shown on the rows in 22_i and are derived from a one-to-one mapping of a subset of 14_j. The elements corresponding to the packaging format are shown in 22_j mapped from 16_j, and the child outcomes presumed to be effected through the regulating variables as they are mediated by the teacher are mapped from 14_k, or 15_k, or 16_k depending on the stage of development.

Evaluation of Anthropoactive Products

Based on this conception of anthropoactive products, it is clear that from an evaluation point of view the evaluation of a training program is a two-step process. At its simplest level of analysis, evaluation of the

Figure 22. . Characterization of anthropoactive products



packaging format of a training program is focused upon the degree to which the training program produces the desired level of the several regulating variables operating in or through the persons trained by it; while evaluation of the effects of the several regulating variables is focused upon the multivariate set child outcomes which each is presumed to be causally related to, although it may be difficult, in practice, to disentangle the separate effects of each. A possible evaluation sequence is shown in Figure 23.

Insert Figure 23 about here

The degree of implementation discussed earlier in the paper reflects some of these considerations. A somewhat more complicated but much stronger model or conception of the evaluation design would require specification of teacher mediated variables (T), variables mediated through curriculum materials (Z), and any of their interactions (TZ); and at least three general classes of child outcomes: learning-to-learn or mathemagenic outcomes (Y), cognitive, achievement, or related outcomes (Y'), and various kinds of attitudinal or other outcomes (Y''). The model might take the following general form:

$$Y + Y' + Y'' = f(T + Z + TZ)$$

Since major relationships between the vectors of the matrix of independent

Figure 23. Minimal characterizations required for evaluation of anthropoactive products

Components of Training Program	Desired Level of Regulating Variables Via Teacher Behaviors	Child Outcomes
C_1	X_1	Y_1
C_2	X_2	Y_2
C_3	X_3	Y_3
.	.	.
.	.	.
C_N	X_N	Y_N

Focus of Evaluation of Packaging Format for training program

Focus of Evaluation of Regulating Variables achieved through training program

Note: Each component of the training program is presumed to be included because it is related to the attainment of one or more described teacher behaviors which are in turn presumed to be desirable because they are related to one or more desired child outcomes.

variables and the vectors of the matrix of dependent variables (if hypothesized prior to the data collection) might be directly testable, the model should be stronger than the simple, non-specific program or no-program type of evaluation design often used. David Wiley (1971) has advocated the use of even more complicated program evaluation models which include home related variables (H) and child characteristics (C) as terms in the model, to more adequately represent the complete set of events which might be causally related to a hypothetical set of child development outcomes of interest to program developers. Obviously, such models require far more careful evaluation planning, design, and analysis than most conceptions of program evaluation have previously considered.

Information Products

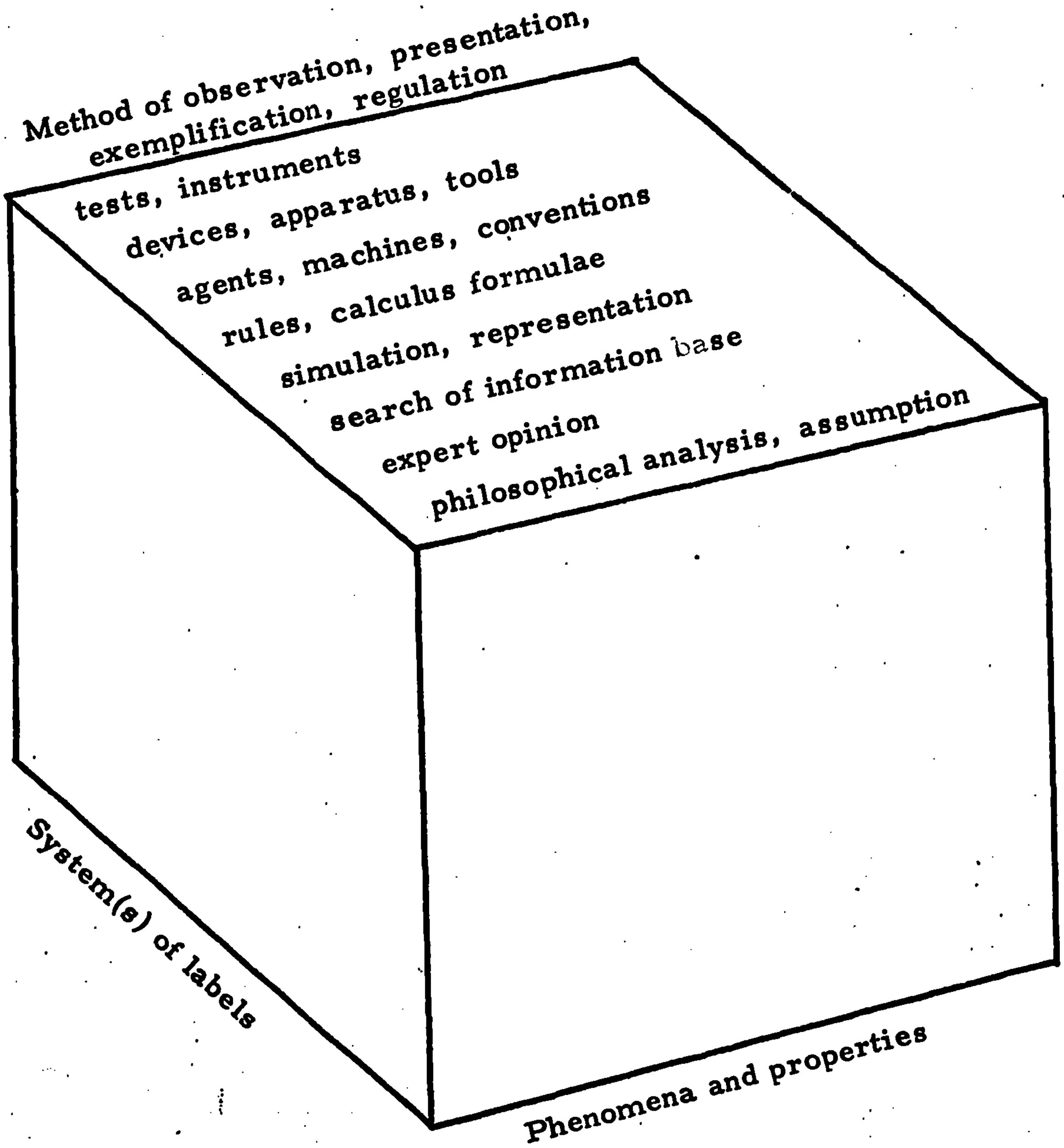
In the later sections of the paper a reference was also made to the several types of knowledge products which emerge from the R&D process in the physical and in the social and behavioral sciences. These knowledge or information products as we will subsequently be referred to are essentially of two kinds (a) scientific information products which help to constitute the knowledge base in the sciences, i.e., the research reports, technical documents, and manuscripts, etc., which reflect the results of the basic and applied research, and (b) pedagogic information products which seek to convey the scientific information in a less technical form to those who want, need or require its assimilation, i.e., the text books, curriculum materials, lessons, etc., which form a large part of the set of products derived from the educational research and development effort.

The results of the finding of the basic and applied research stages in the physical, social and behavioral sciences (and in other sciences as well) are usually transformed into these scientific products as illustrated in the matrix in Figure 24. The results of scientific observation are first

Insert Figure 24 about here

transformed and transmitted by means of some alphabet or other symbol system to become a set or system of labels each of which references at least some important properties, relations, attributes, or attribute-values of the phenomena which were observed. These labels constitute the elements of the rows for the characterization of the scientific products which form the knowledge base. The method of exemplification, or regulation of the phenomena and its properties is mapped from 9_j or 13_j onto 24_j and the phenomena and its properties are mapped from 9_{ik} or 13_{ik} onto 24_{ik}. Thus, the labels are the terms invented by the scientist to reference the phenomena of the world and their properties as the language system for discussion; these phenomena and properties form the definition for the labels and the methodology of regulation become the means to exemplify

Figure 24. Characterization of the formal knowledge base for the physical, social, behavioral, and other sciences



the "reality" to which the labels apply.

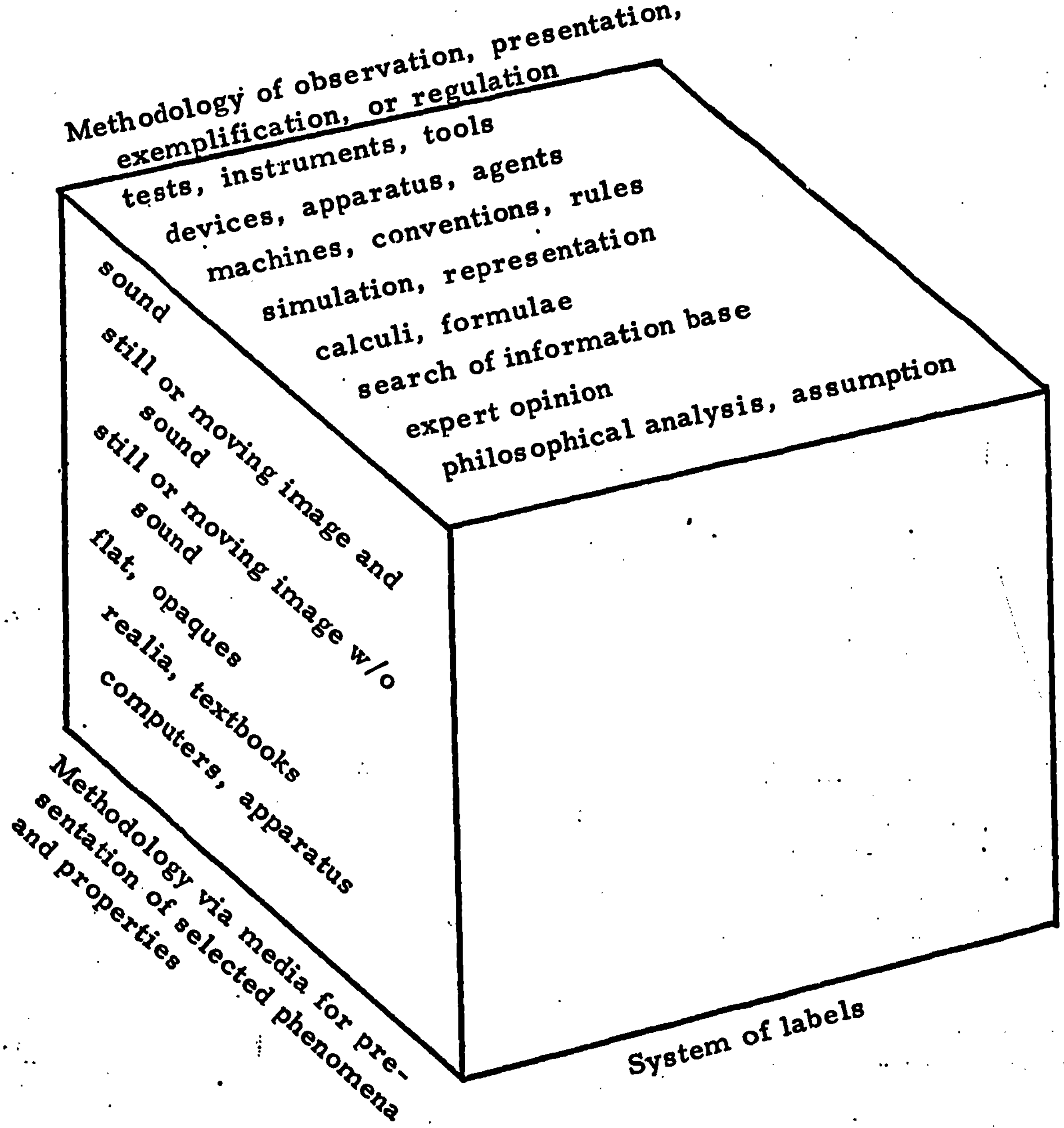
Pedagogic Products

The last type of product to be analyzed are those we are most familiar with, and which form the major justification for the educational research and development activities currently under way in laboratories and R&D centers. These are the pedagogic products, the games, toys, books, slides, films, etc., which constitute the means for transmitting subsets of the scientific knowledge base to others who cannot assimilate this "knowledge" in its "scientific" form for a variety of valid or acceptable reasons. The minimal definition matrix for the characterization of these products is shown in Figure 25. The system or set of labels for the

Insert Figure 25 about here

pedagogic product in 25_i are derived from a one-to-one mapping from 24_i although they may need to be mediated by another alphabet or symbol system depending on the relationship between language level of the user and the language of instruction. The methodology for exemplification or regulation of the phenomenom is shown as 25_j and it is mapped from 24_j. The media for the presentation of the phenomena and its properties as defined by the methodology of exemplification represent the several forms that pedagogically relevant phenomena might be "captured" or "displayed" in.

Figure 25. Characterization of content of pedagogical products



Product Objectives and the Choice of Content for Pedagogic Products

Up to this point, we have tried to present a general model of the educational R&D enterprise and of the products that emerge from it. The decision to engage in programmatic R&D in certain areas of education forces the developer (or the sponsor) into setting priorities, and into choosing a limited set of objectives for which products will be developed. The specification of these objectives is often done at several levels of abstractions removed from the operational form that they take via the pedagogic product route. These objectives whatever their level of specificity are the focus of the evaluation activities of interest to the product developer, and it is here that the "formative" evaluation activities discussed by Scriven (1967) appear to serve their major purpose.

A minimal definition matrix corresponding to a hypothetical set of general objectives for the child to be derived from the use of pedagogic products as they relate to the knowledge base in 24_{ijk} is depicted in Figure 26. On the rows, objectives related to the production or reproduction

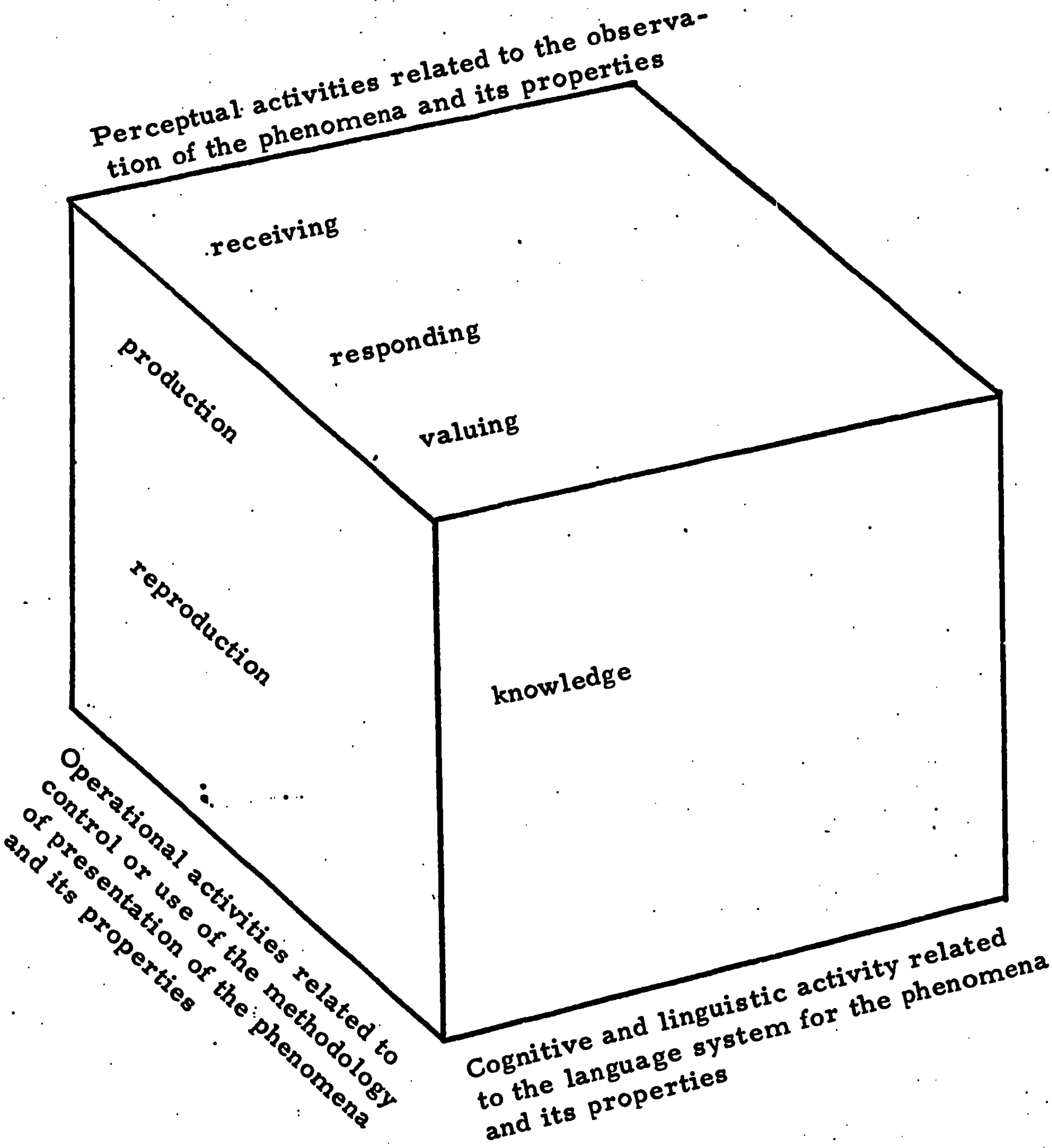
---

Insert Figure 26 about here

---

of the methodology of exemplification and regulation of the phenomena and its properties are depicted mapped from 24_j, and corresponding to the potential vocational or efficiency outcomes desired by some educators. On

Figure 26. General objectives for education and for pedagogical products used in education



the columns, objectives related to the observation and perception of the phenomenon and its properties are depicted mapped from 24_k and intended to reflect the sensory development or "feeling" outcomes desired by some educators. On the blocks, objectives related to the facility in the characterization and labeling of the phenomena are depicted mapped from 24_i, and intended to reflect the communication and "knowledge" outcomes desired by other educators.

Obviously the matrix of objectives displayed in Figure 26 is at too high a level of generality to be of any value to the development group in arriving at a set of discrete pedagogical products. But it does reflect the orientations and biases that might accompany the development effort.

Thus, a more "specific" set of objectives is required as a basis for content choice which will need to be made prior to or concurrently with the R&D effort.* Such a set is displayed in matrix form in Figure 27.

---

*Three broad classes of educational strategies can be conceived as having application to the development of pedagogical products to a given area of the knowledge base related to (a) the source of the structure of the knowledge base, and (b) the verificational or validational procedures to be employed by the user. These will be referred to as the <u>Didactic</u>, <u>Discovery</u>, and <u>Experience</u> models.

In a <u>Didactic</u> model the structure of the knowledge base as reflected in the concepts, definitions, and propositions of the language system for it is derived from some external source, and the verificational procedures for the individual are handled axiomatically. Thus, the system of labels or language system directly becomes the conceptual system of the individual, as in the classical approach to concept learning.

[footnote continued on page 103]

Insert Figure 27 about here

On the rows, the phenomenal domains, the properties to be explicated and their appropriate labels, and the requisite methodology for exemplification of the phenomena are presented mapped from a subset of 24_i, 24_j, and 24_k of the knowledge base. On the columns, the modes of perception to be

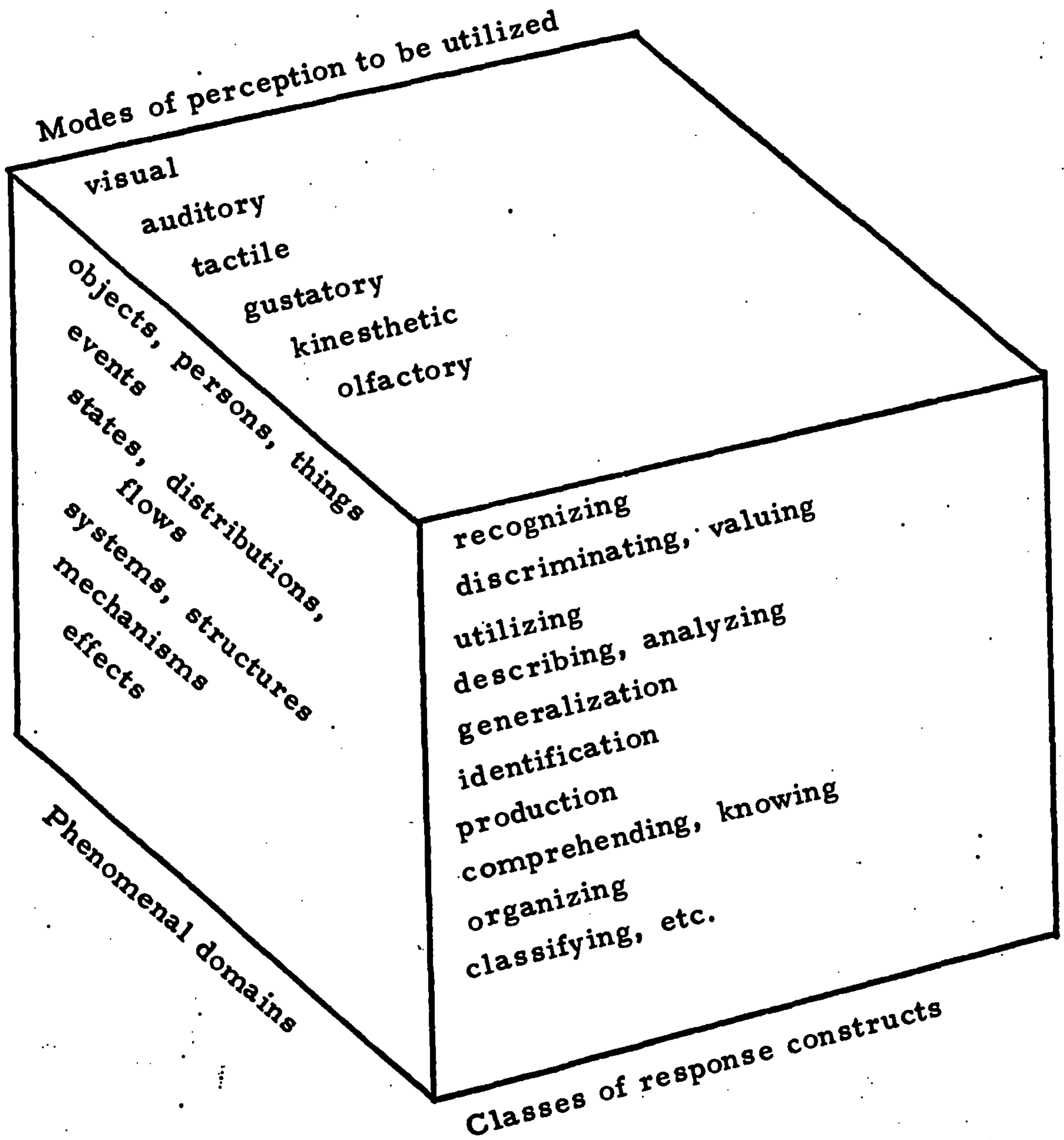
In a <u>Discovery</u> or empirical model, the structure of the knowledge base as reflected in the concepts, definitions, and propositions of the language system are derived from some external source, but the verificational procedures for the individual are derived from the variables of immediate experience with the phenomena. Thus the language system for the knowledge base is indirectly mediated by the quality of the experience before it becomes the conceptual system of the individual.

In an <u>Experience</u> or existential model, no structure of the knowledge base is imposed from any external source. The source of the knowledge base and the verificational procedures for the individual are both derived from varieties of immediate experience with the phenomena. Thus the conceptual system of the individual becomes an acceptable language system for the phenomena of the knowledge base.

It should be pointed out that the discussion within the paper has implicitly assumed that a discovery model could and would be utilized as the basis for development of educational products for almost all subject matter areas.

Each educational model presents certain unique difficulties from an instructional viewpoint. Major problems with the Didactic model arise when immediate experience is not consistent with the language system and the axioms are in conflict with experience as the verificational source. Major problems with the Discovery model arise when immediate experience is not appropriate to the language system. Major problems with the Experience model arise because immediate experiences are always restricted to a very narrow range by operational, physiological, psychological, and environmental constraints. (Thus structure is imposed anyway.)

Figure 27. Specific objectives for pedagogical products



utilized are presented mapped from 25_j and reflecting any derivatives from the knowledge base in the social and behavioral sciences. On the blocks the response constructs of interest are presented mapped from a subset of 14_k.

In Figure 25 we have presented a definition of the product in terms of its content, i. e., the requisite operational form of the objectives. In Figure 27, we have presented the developers' bases for choice of content of pedagogical products, and in Figure 26 we have presented a set of general objectives for which a sequence of pedagogical products might be chosen. In effect, the form of the products depicted in 25_{ijk}, is the direct or indirect operationalization of the set of specific objectives of 27_{ijk}, and the general ones of 26_{ijk}. Scriven's carefully reasoned concerns over the evaluation of the goals of the developer is specifically focused on the all too general lack of congruence between the elements of these three matrices. This continued concern is essential to the eventual explication of the true product characteristics by the evaluator/developer <u>prior</u> to the final packaging and testing.

A representation of the hypothetical flow of a pedagogical product between the applied and development stages as mapped through the series of minimal definition matrices previously depicted in Figures 24 to 27 is shown in Figure 27a. In the figure the product form is seen as undergoing

Insert Figure 27a about here

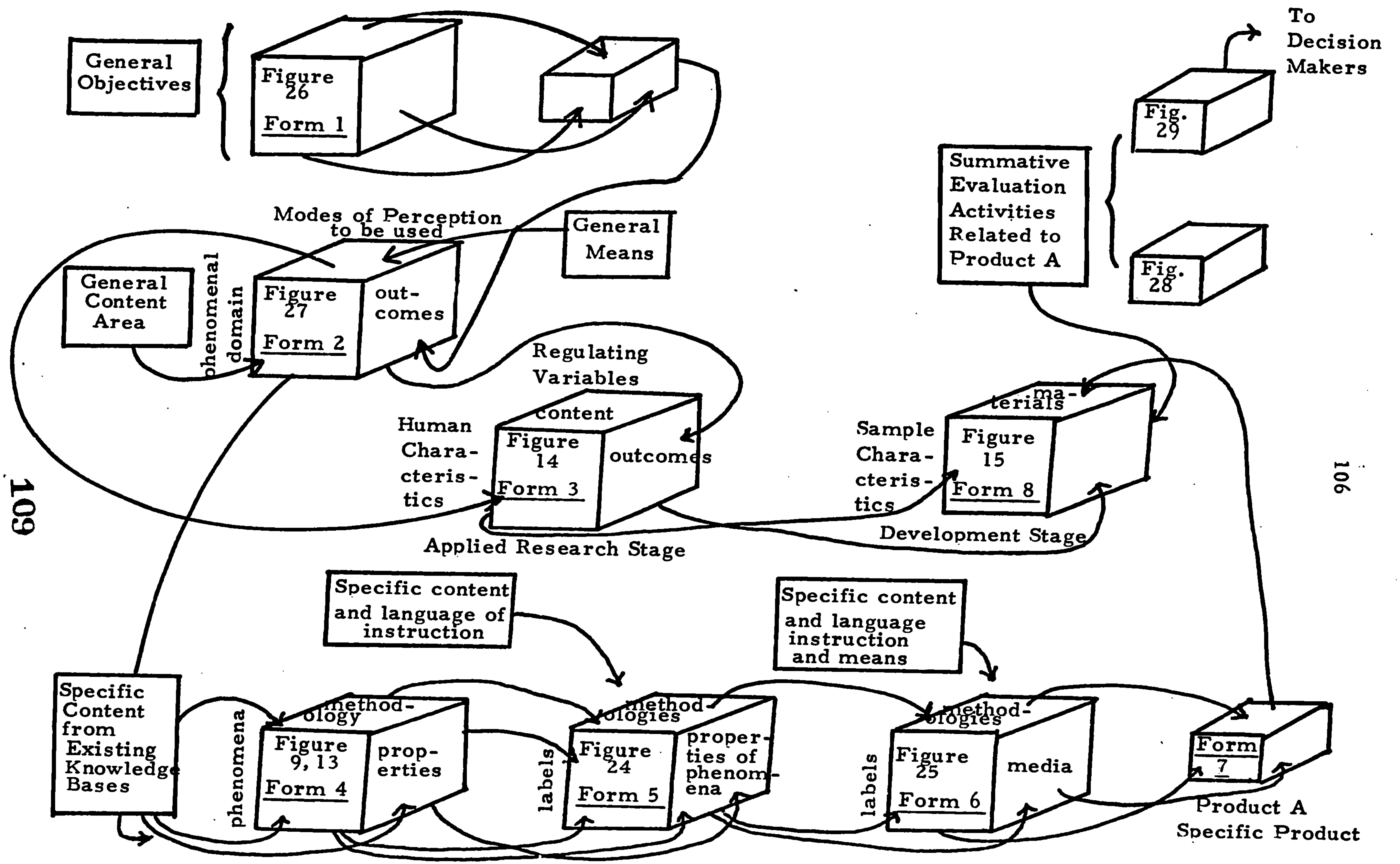


Figure 27a. Hypothetical flow of pedagogical products between applied research and development stages to achieve mapping from content variable to educational materials.

a series of mappings or transformations of its elements culminating as one cell of the minimal definition matrix for the development stage of the social and behavioral sciences outlined in Figure 15. The flow begins with Form 1 which is shown in Figure 26 where the General objective for the product are agreed upon which force choices among outcomes. Form 2 is shown in Figure 27 where the General content area and the General means are agreed upon to attain the outcomes mapped from Figure 26_{ijk}. Form 3 is shown as Figure 14 where the product is defined and locatable within the R&D continuum at the applied research stage, as one of a set of undifferentiated content variables 14_j, to achieve a set of outcomes 14_k, for a group with certain characteristics 14_i. Form 4 of the product is where the specific content is selected. Form 4 is knowledge base created by the information obtained through basic research in the physical, social, behavioral (and other sciences as well) as depicted in Figures 9 and 13. Form 5 is shown in Figure 24 where the specific language of instruction is added to the specific content mapped from the previous form. Form 6 is shown in Figure 25 where the specific means of exemplification has been added to the specific content and the specific language of instruction mapped from product form 5. Form 7 is shown as a specific product, Product A, manufactured or constructed in a prototype version and mapped from all of the previous product forms denoted. Form 8 is shown as Figure 15 where the product is now defined and locatable within the R&D continuum at the development stage as one of a differentiated set of products to achieve a set of specific outcomes for a sample with a specified set of characteristics.

Using this hypothetical product flow as a framework for viewing the types of evaluation activities that may be required, it is apparent that a focus on the adequacy of the mapping is one of the more important roles of the evaluator as Scriven has suggested with his concern over the "formative" evaluation issues. However, formal evaluation of product A is most often done when the product is in Form 7, and this evaluation is concerned with more precise specification of sample characteristics and outcomes for the given product to characterize it either in Form 8 and earlier in Form 2. The minimal definition matrices for the evaluation data, and the evaluation data analysis also shown in Figure 27a represent the procedures used by the evaluator in arriving at information to provide some of the decision makers listed in the first part of the paper. Thus, the evaluator's major activities are seen as focusing on (a) the adequacy of the mapping, or (b) the degree of specificity of information in these minimal definition matrices.

Summative Evaluation

The focus and format of the data for the "summative" evaluation of these pedagogic products is shown in data matrix in Figure 28. Elements on the rows correspond to a specification of the response methodologies

Insert Figure 28 about here

normally used in the social and behavioral sciences for the data on pupil outcomes, including identification, pair or n-item comparison, recognition, matching, production, reproduction, estimation, categorization, ranking, observation, etc. The types of criterion used in evaluating or judging the data from evaluation studies on the columns and these include norm-based, criterion based, and domain referenced criteria with the referent either being the item forms, responses, or the knowledge base (and any theory or models as criteria), and descriptive or non-criterion studies. The response constructs on the blocks of the data matrix are mapped from 15_k, and are assumed to be representational but fallible indices of the objectives specified in 27_k. The evaluation data defined by the criteria model can be mapped onto the elements of a data analysis matrix illustrated in Figure 29. In this data matrix, the elements of the rows corresponding to a set of experimental designs for controlling the

Insert Figure 29 about here

interpretability of the data, the colums a set of measurement scales for transforming the data to a set of numbers, and the blocks a set of statistical procedures for determining the goodness of fit of the data to the criterion in 28_j.

Figure 28. Focus and format of summative evaluation data

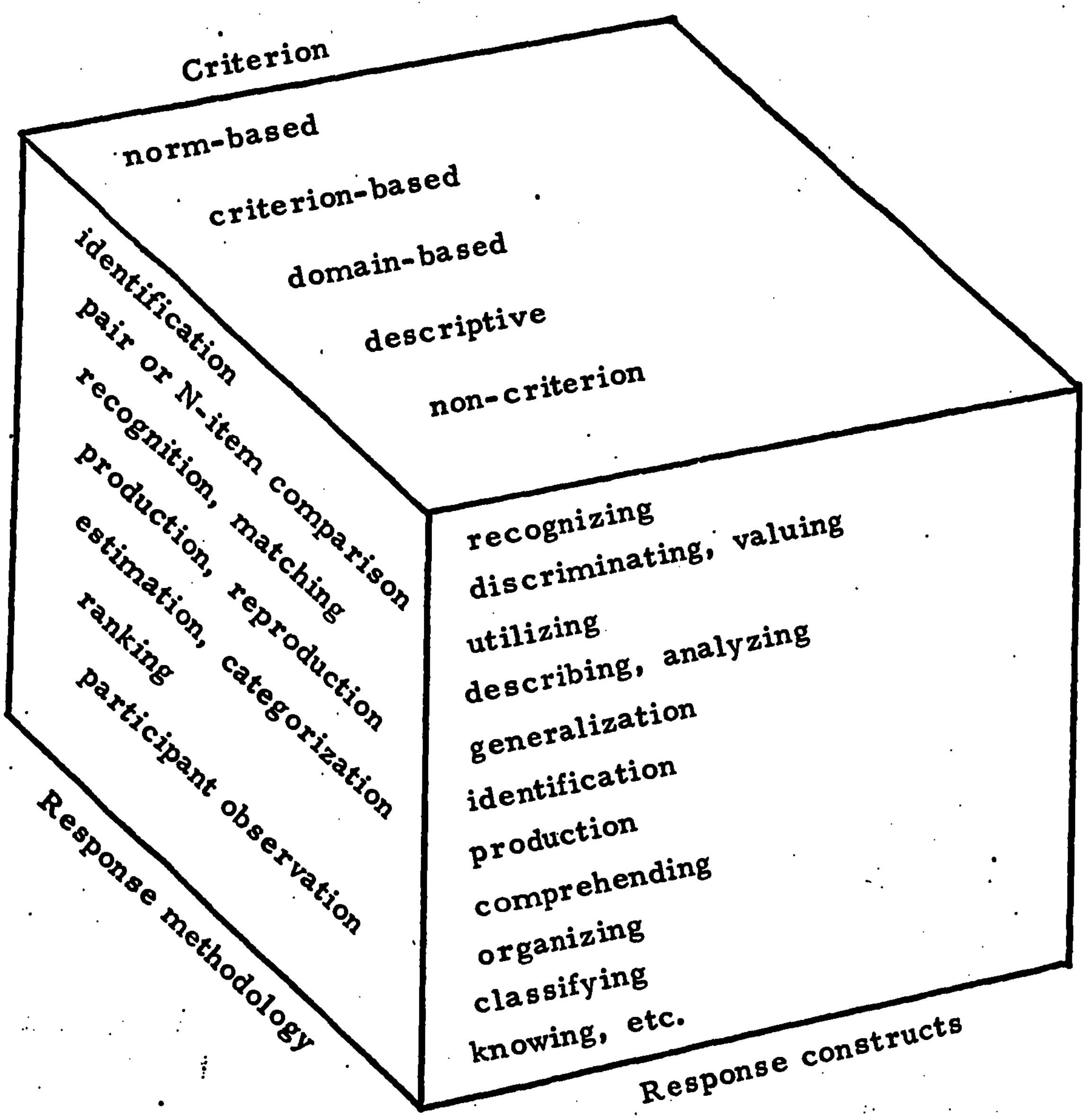
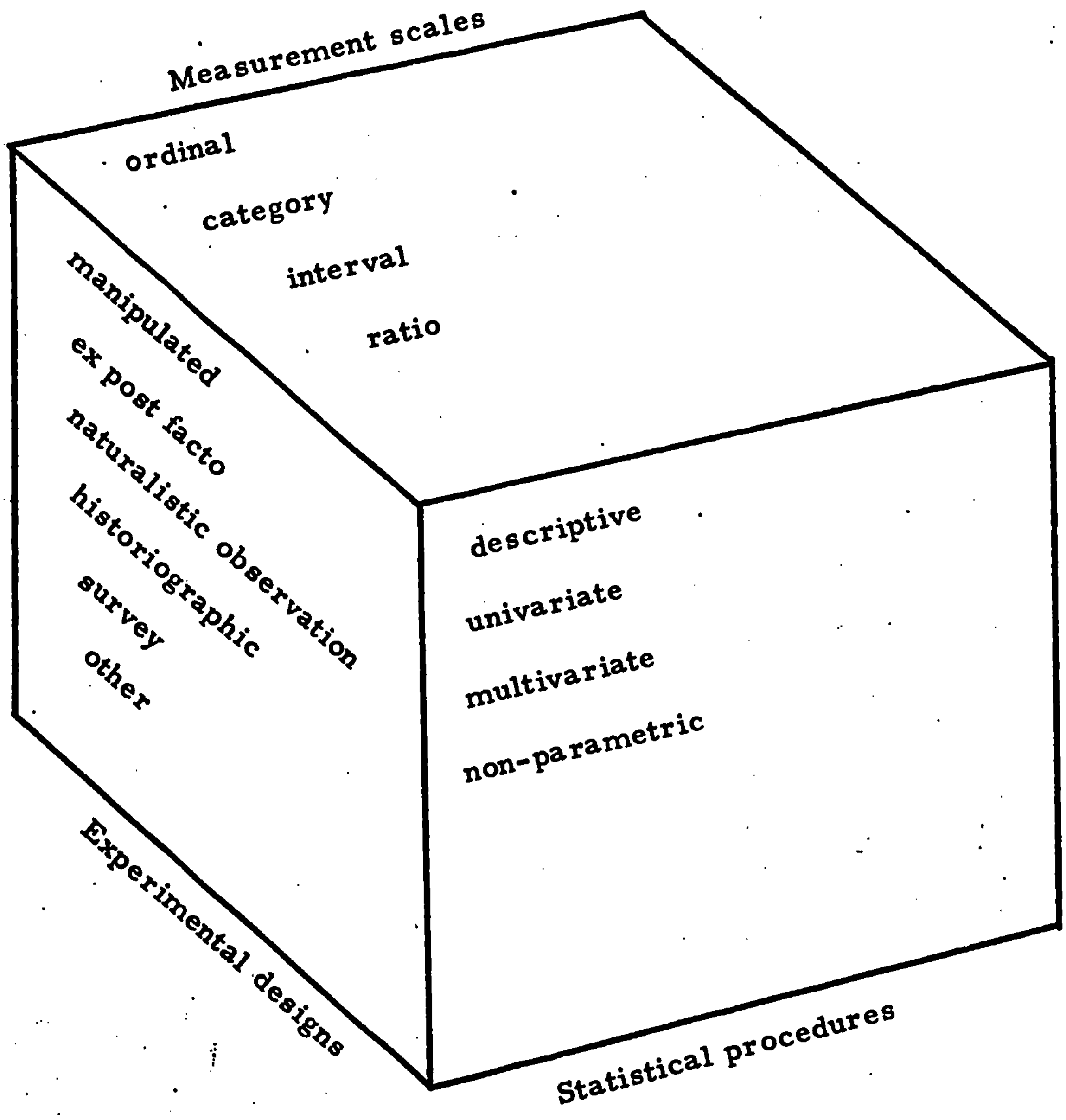


Figure 29. Focus and format of summative evaluation data analysis



Summative evaluation activities which are undertaken to provide information to skeptics, to other evaluators, to cooperating R&D institutions, and to some types of consumers are usually focused on the elements of these two matrices.

However, many consumers will be equally interested in a more extensive set of information related to one or more of the other elements of product characteristics, performance specification, and to system requirements as shown in Figure 6 (page 30), although they would use criteria of the same kind noted on 27_j.

In the discussion of formative and summative evaluation activities, we have tried to tie the information interests of each of the types of decision-makers discussed earlier to the particular phases of the R&D process, and to the product forms that emerge at that level. The only group not mentioned thus far is the major sponsor of most of the R&D activities under way in labs and R&D centers, the federal government.

Evaluation for Policy-Decision at the Federal Level

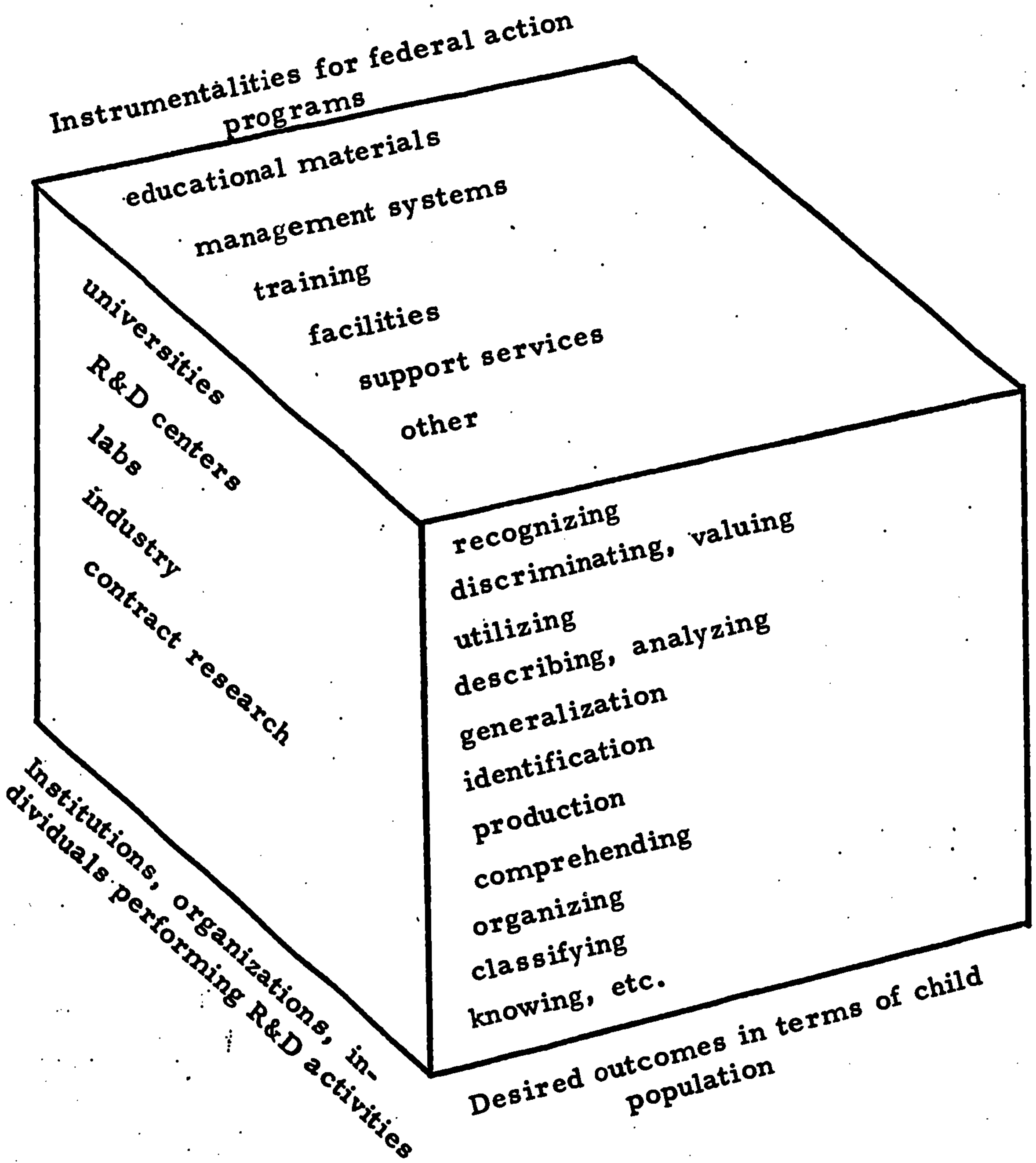
The USOE is the primary federal agency sponsoring educational research and development work represented by the labs and R&D centers. As part of their contracts with these organizations, OE normally requires submission of all evaluation reports developed by the contracting organization. As the steady flow of these reports wend their way to Washington, the receipt of such reports may tend to promote a sense of assurance on the part of many federal planners and developers that they (OE) are receiving evaluation information which is useful for the policy decisions they might face. A macro structural framework for viewing the educational R&D enterprise from the federal level, and a possible framework for their evaluation of it is depicted in the Figures 30 and 31.

Insert Figure 30 about here

The desired outcomes for a target population are shown on 30_k mapped from 15_k, as determined or selected by some type of formal assessment procedure or from manifest or assumed needs detected or mandated by the Congressional, or Executive, or Judicial branches of Government.

The columns 30_j show the set of instrumentalities selected for the attainment of outcomes related to the solutions to these needs which are mapped from 19_j, and thought to be a feasible bases for operational action programs. The rows show the organizations and individuals who are contracted with to perform the R&D activities presumed to be required for the production of the instrumentalities or technological solutions mapped from 17_i, or a variant set of them. Precise specification of many federal programs or projects can be derived by performing a more extended differentiation of desired outcomes according to (a) certain types of populations,

Figure 30. Overview of federal R&D activities in education



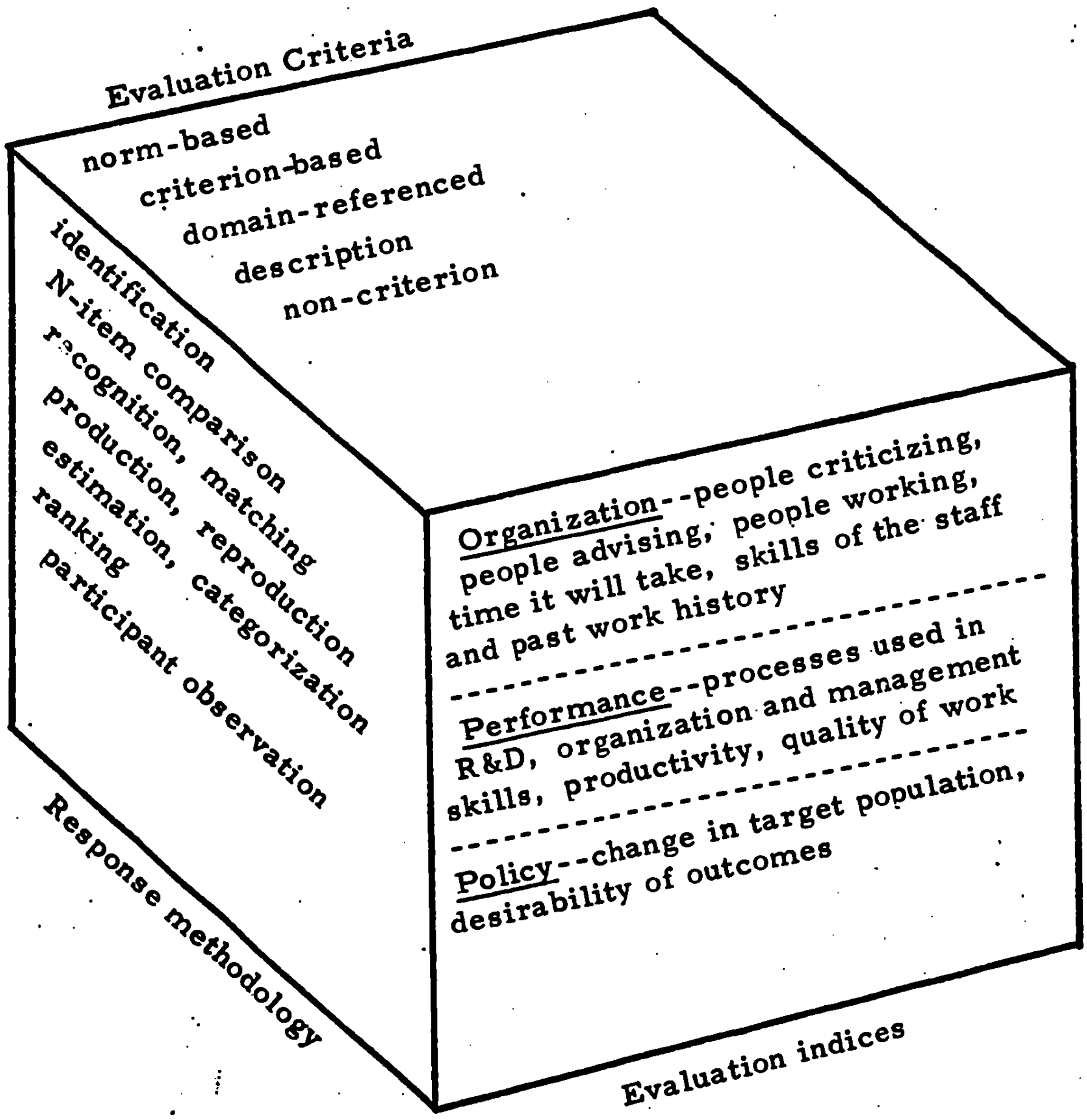
i. e., disadvantaged, handicapped, early childhood, (b) certain types of location characteristics, i. e., urban, Appalachian, etc., or (c) certain types of outcome derived classifications, i. e., vocational, reading, science, etc. In other instances precise specification can be achieved by (d) greater differentiation of the instrumentalities, i. e., curriculum, libraries, laboratories, day care centers, etc., and (e) extended differentiation of the types of R&D or activities, i. e., basic research, development, training, diffusion, etc., or the organizations and individuals who work on them.

The means of achieving more precise definition of the federal project is important for examining the focus of evaluation of all R&D programs at the federal level as is depicted in Figure 31. The rows show the

Insert Figure 31 about here

response methodologies required for the generation of data at this level achieved by a one-to-one mapping from 25_i. The columns show the evaluation criteria used at this level, and they are achieved by a one-to-one mapping from 25_j. Three types of evaluation indices often used in federal evaluation efforts are shown on 31_k. Performance indices are used to evaluate the production of the products, materials, or packages, the quality of the products or materials, the time they take to produce the processes used, etc., etc. Organizational indices are used to evaluate the qualifications of the people working on the project, or advising it (or criticizing it),

Figure 31. Critical factors in evaluation of R&D programs at the federal level



the organizational and management skills they possess, or the reputations and political strengths of the institutions or consortia. Planning or policy indices are used to evaluate the achievement of desired outcomes in the target population. In the final analyses the policy indices are perhaps the more important consideration at the federal level.

Summary

This paper, as the title indicates, has been focused on evaluation in the context of educational R&D. The first half of the paper dealt primarily with evaluation activities, issues, problems, and the process of R&D; while the second half dealt with the context of educational R&D in terms of the products which are derived from it. The presentation of the total scope of educational R&D and its various sub-components, as each might be viewed by one or more of the several decision-makers who have legitimate interests in it helps to provide a common perspective for understanding each others concerns, and a means of facilitating the focus of inquiry or communication. Expanded, more elaborated, or alternative formulations of the totality of educational research and development at the national level are a prelude to more efficient use of human and financial resources for the solution of pressing national problems, and to more additivity of efforts. While the individuals, organizations, and institutions who engage in these R&D activities all have a vested interest in maintaining their part of the action, they also should have an intellectual and public interest in whether they are a part of an integrated R&D system, or merely part of an educational R&D game.