DOCUMENT RESUME

ED 113 510 CE 005 031

AUTHOR Martin, James L.
TITLE Criteria for Developing Criteria Sets.
INSTITUTION Minnesota Univ., Minneapolis. School of Public Health.
PUB DATE Aug 75
NOTE 29p.; Product of the University of Minnesota Independent Study for Health Care Trustees

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS *Educational Philosophy; *Evaluation Criteria; Evaluation Methods; Matrices; *Systems Approach

ABSTRACT

Criteria sets are a necessary step in the systematic development of evaluation in education. Evaluation results from the combination of criteria and evidence. There is a need to develop explicit tools for evaluating criteria, similar to those used in evaluating evidence. The formulation of such criteria depends on distinguishing between terms (concepts) and statements (linkages among concepts). For both terms and statements a further distinction must be made between categories (nonvariable classes or attributes) and continua (variable continuous dimensions). Borrowing from the discipline of sociology, an exploration for categorical terms for constructing a systems approach yields the elements: resources, structures, integrations, performances, and outputs. Continuous dimensions borrowed from theory development are scope, parsimony, precision, and accuracy. Combining these dimensions and elements in a four by five solution grid provides both descriptive definitions and a prescriptive framework. Affixing labels to each cell of the grid provides both literal definitions (labels) and theoretical definitions (dimension by element). Dividing the system's elements into their philosophical components yields ontology, epistemology, ethics, aesthetics, and psychology. Crossing these philosophical elements in turn with the five system elements yields a five by five matrix which provides 25 possible elements for the investigation of criteria for criteria sets. (JR)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED 113 510

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATION STRATEGIES

CRITERIA FOR DEVELOPING CRITERIA SETS

TM 005 031

ERIC

CRITERIA FOR DEVELOPING CRITERIA SETS

James L. Martin
Independent Study Program
for Health Care Trustees

August, 1975

189 Frontier Hall
University of Minnesota
Minneapolis, Minnesota, 55455

ERIC

This paper is distributed for general interest. Reproduction in whole or part is permitted if proper credit is given. This distribution neither expresses nor implies approval of its contents by the Program in Hospital and Health Care Administration or the W.K.Kellogg Foundation.

INTRODUCTION

Definitions of Evaluation have consistently of late included criteria or standards as a necessary or essential element (20). Since this definition is commonly accepted as consensual fact, this investigation attempted to build toward a theory of evaluation using this definitional element. This study considered questions as:

How can you recognize a good criterion, or a good set of criteria?

How can you develop a good criteria set?

Where do good criteria sets come from? and,

Who can design good criteria sets?

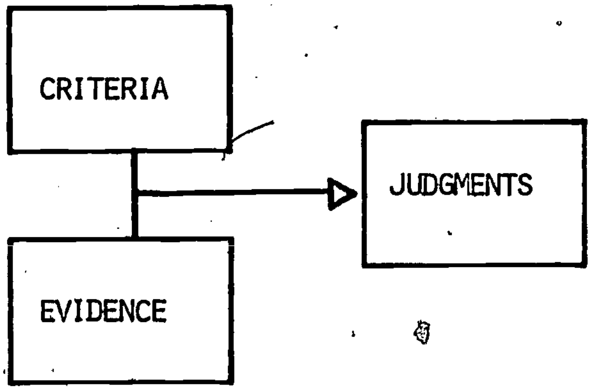
Little is offered in the literature of Evaluation to suggest how criteria are developed, or where they come from, let alone the idea of "goodness". Worthen (21) reports 25 general tasks for evaluation suggested from three different surveys. Of these 25, four can be perceived by this writer as associated with criteria sets. Fourteen seemed most applicable to evidence (measurement and statistics), six to judgments and four, other activities: disseminating, recommending, providing feedback and managing resources. Of the 81 competencies needed to perform these 25 general tasks, 10 fell within the area of criteria, 51 within evidence and 6 within judgments. Fourteen lay within other areas. This suggests either a lack of emphasis of criteria or possibly criteria are less important. Because the definition of criteria-compared-to-evidence implies equal importance, this investigator assumed criteria had been underemphasized in the past. This is supported by research for, but presently inadequate, theories of setting performance criteria in mastery education (2), an area currently popular in education. The objective of this investigation was to: use fruitful descriptions of evaluation as a base to explore, explicate and embed within a framework, criteria for criteria sets.

Criteria for criteria sets, then, are viewed here as a necessary step for further systematic development of evaluations in general, and within Adult Education in particular.

ERIC

EVALUATION AS A UNIT FOR INVESTIGATION

Evaluation has evolved definitionally from an emphasis on measurement to: Criteria and Evidence which lead to a Judgment. Pictorially this is viewed as:



These three "boxes" are viewed as elements of a single system--Evaluation. These three elements provide a descriptive definition, but because they are discrete classes, provide only a partial set of tools to perform analyses, explanations, and predictions. To move beyond description (a fruitful and necessary stage) continuous terms and statements are also necessary (6, p. 173). This point of necessity raised the questions of sufficiency as well. This investigation was not only interested in the necessary conditions for developing a theory of evaluation (of which high quality criteria are only a part), but also in the sufficient conditions. Hage (6, p. 172) has suggested a set of necessary and sufficient conditions for theory development in Sociology:

> . . . a theory should contain not only concepts and statements but definitions--both theoretical and operational--and linkages, again both theoretical and operational.

To date this set of conditions for theory development has been extremely fruitful when applied by this investigator to evaluation systems (11) and to educational systems (12, 13).

ERIC

This set of six conditions (concepts, statements, theoretical definitions, operational definitions, theoretical linkages and operational linkages), suggests a design that explores and explicates each of the elements in this necessary and sufficient set. This would appear as the next step in developing a theory of evaluation.

Current evaluations, as viewed by this writer, are concerned with only a subset of the system elements suggested here. This investigation assumes that a continuously productive approach hinges on a more effective design strategy--one that considers a complete set of necessary and sufficient conditions.

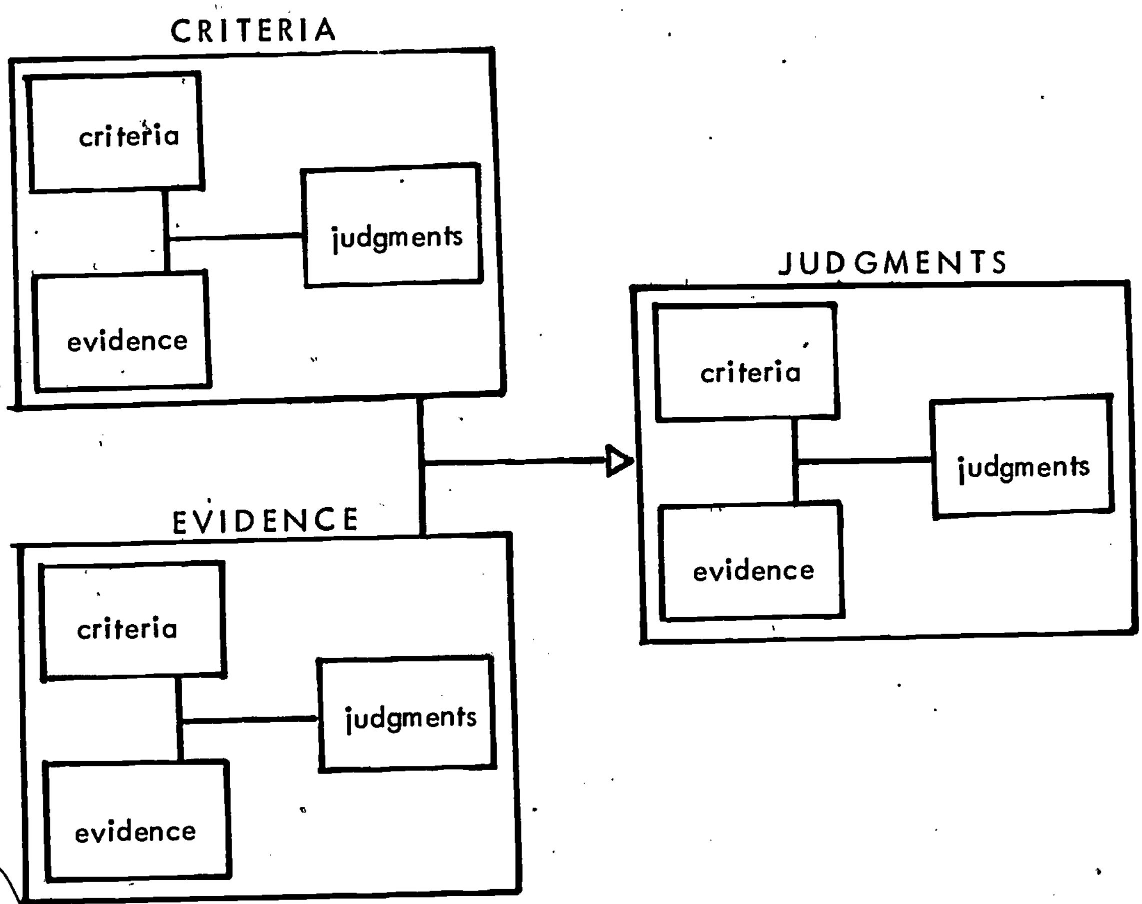
STRATEGIES DESIGNED TO DEVELOP BROAD PERSPECTIVES

The strategies discussed here have been gleaned from Hage (6, 8), Nadler (17, 18), Dubin (3) and Johnston (9). They have been previously applied successfully to evaluation (11) and open learning systems (12, 13).

The basic design strategy is to first design the formulation intuitively or based on whatever scraps of data are at hand--your first guess-timate is always a free one.

Second, explicate the underlying framework. Third, check evidence with criteria, and finally, concentrate on tautologies and inconsistencies within the developing framework.

The first insight provided by this strategy is that each of the three elements in the definition (criteria, evidence, and judgments) contains criteria, evidence and judgments at a more specific level. Pictorially this is displayed as:



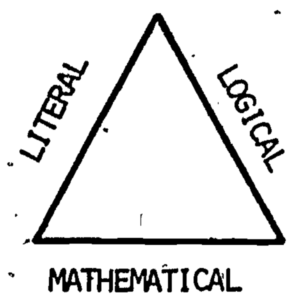
Within the Evidence box, criteria of evidence (e.g. validity and reliability), when compared against the evidence about evidence, allow for conclusions about the evidence at hand. In short, one evaluates the evidence. This is done much of the time--explicitly and precisely. Likewise it is assumed Criteria, as a unit of investigation, can be evaluated, as can conclusions or Judgments. Neither of these investigations seem to be performed explicitly at present.

Because of a lack of explicit tools to evaluate criteria, and because Criteria is seen as an important element of evaluation, the remainder of this investigation concentrated on Criteria as the unit of inquiry or investigation. An explicit set of tools to evaluate Criteria would balance-out the wealth of explicit tools presently available and applied to evaluating Evidence. This balance would then allow for developing Conclusions systematically.

A second insight was that diversity (preferably by design) is needed to develop broad perspective. In application, this diversity was most fruitfully supplied by the diverse languages of logic, mathematics and literature. These distinctions are explicitly laid out in the Economic Education Bulletin (4, p. 14-15).

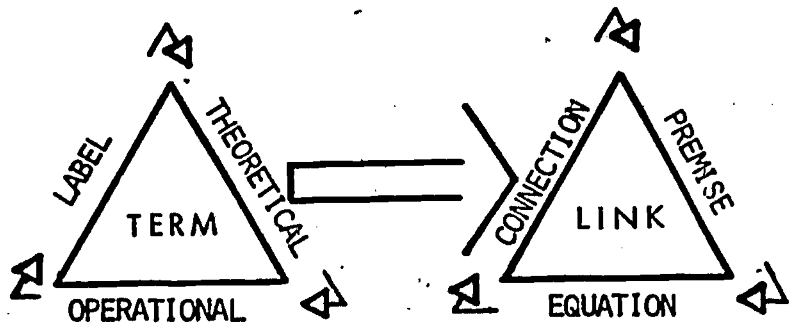
> Logic...has been developed by analyzing...actual conduct of inquiries... Application of this theory to any inquiry ordinarily facilitates progress.
>
> Mathematics...serves not only as a means of analyzing the connections...but also as a means of suggesting other possible connections, and therefore of directing the inquiry to new data or connections not previously investigated.
>
> The language (verbal) tool is used as a means of describing ...when properly used, the language tool makes possible the construction of a word picture that correctly represents the steps in the inquiry.

In this sense, the elements of the necessary and sufficient set form a triad of logical, mathematical and literal.



ERIC

Distinguishing between Concepts and Statements (linkages among concepts) while retaining each of the three languages explicates the necessary and sufficient elements as:



This is a more specific statement of the more common definition of a theory: concepts linked together.

These six elements form the necessary and sufficient set based upon the definition of theory development stated previously and applied to this inquiry:

Each of the six has a specific purpose or function.

Literal terms (labels)	provide	description
Logical terms (???)	provide	meaning
Mathematical terms (???)	provide	measurement
Literal links (connections)	provide	analysis
Logical links (premises)	provide	plausibility
Mathematical links (equations)	provide	testability

The arrows indicate the style of search that has proven most effective for this investigator. By definition, Concepts are necessary to form Statements. Concepts act as Terms. Then, linked together they form Statements. Given labels, search for meanings. At this point one can skip to verbal connections and premises without searching for measures. What isn't explicated here is how one searches for Terms when starting with Statements. To date the scarcity of statements in literatures of both evaluation and open learning has prevented the leisure of starting with Statements.

Distinguishing between concepts and statements is the most common definition of theory. This definition has been extended here to suggest 3 parts for each concept and 3 parts for each statement.

Another distinction is implied with this explication of six elements. Some examples within each of the six are categorical classes or attributes. Other examples are continuous dimensions or variables. Categories are non-variable. Continua are variable.

Adding the distinction between categories and continua to the distinction between Terms and Statements leads to a fourfold table:

	TERMS	LINKAGES
CATEGORIES		
CONTINUA		

This allows for a third insight. No one cell is more important than any other. All four are needed to satisfy this definition of theory development. Thus, all four are needed to develop a theory of evaluation which includes criteria for criteria sets. With these four possibilities explicated a fourth insight suggested that each cell requires techniques to:

Explore or Search,
Explicate or specify, and
Embed or structure

In addition, ways are needed for moving from cell to cell, or understanding how the four relate together as a single system (cf. 14).

Before continuing, an explication of uses for each cell may help clarify subsequent applications.

> Categorical terms are used for defining units of inquiry, when identifying cutpoints or qualities and when identifying elements or major classes of phenomena.

ERIC

Continuous terms are needed for Theoretical Definitions and Operational Definitions. This is crucial. All derived terms are continuous. These are the building blocks of each continuous statement.

Categorical statements are mandatory for definitional premises and for thresholds along operational linkages.

Continuous statements are needed for Theoretical Linkages and for Operational Linkages.

These four cells suggest productive uses for each. They do not consider categorical links among continuous terms nor do they consider continuous links among categorical terms. They do not mix categorical and continuous modes. In general, these mixed modes are unproductive.

EXPLORING, EXPLICATING AND EMBEDDING CRITERIA FOR CRITERIA SETS

Conceptual Definitions: Literal, Logical, Mathematical

The unit of inquiry is Criteria of Criteria as an element of evaluation. This inquiry is delimited to one of the nine elements specified in the specific definition of evaluation. At this point Criteria of Criteria remains a black box--as yet undescribed, meaningless and not measured.

According to the first cell, in the explicated fourfold table, exploring for categorical terms should yield major classes of phenomena. Initial explorations were directed toward a systems approach (it had been fruitful in past inquiries). Consequently elements of a system were borrowed from Sociology (6, p. 234). This is also consistent with Kreitlow's advice of borrowing from a related field (10). The elements borrowed were Resources, Structures, Integrations, Performances, and Outputs. Synonyms for these elements are Inputs, Arrangements, Articulations, Adequacies and Accumulations.

These elements provided major classes of possible phenomena operating within the investigative unit of Criteria of Criteria. So far only labels have been suggested.

Continuous terms are needed for specifying theoretical definitions, which provide meaning; and operational definitions which provide measurement. Again dimensions--variable but contentless--were borrowed from theory development (6, p. 178) and evaluation (11). These dimensions are Scope, Parsimony, Precision and Accuracy.

As originally applied to theory (6, p. 178-180), Scope was operationalized as a ratio of derived terms to primitive terms or: $\frac{\text{derived terms}}{\text{primitive terms}}$.

In other words, given a certain number of primitive terms the maximum number of possible derived terms can be determined using a mathematical formula, since derived terms are some combination of the primitives. The derived terms in the $\frac{\text{derived}}{\text{primitive}}$ ratio have a defined upper bound. This is consistent with the idea that all inductive conclusions are in some sense probable. A priori mathematical probability is defined as the ratio of $\frac{\text{favorable outcomes}}{\text{possible outcomes}}$. This discussion suggests that derived terms are simply the ratio of $\frac{\text{favorable}}{\text{possible}}$ or defined by an a priori mathematical probability. This is suggested to explicate the practice of including only favorable outcomes in one's logical derivations.

By combining the idea of a priori mathematical probability (favorable:possible) for derived terms with the idea of derived:primitive as a measure of Scope the explicit operational definition of Scope becomes:

$$\frac{\frac{\text{favorable}}{\text{possible}}}{\text{primitive terms}}.$$

This definition indicates that the wider the ratio, the more general the Scope. This ratio transfers Scope as a dimension out of the domain of raw numbers or amount and into the domain of relational units.

The theoretical definition of Parsimony is explaining as much as possible with as little as possible. Parsimony is operationally defined as a ratio of: $\frac{\text{equations}}{\text{premises}}$. There is a large but finite set of equations. Again, only equations which offer explanation (a favorable outcome) are retained. Therefore the explicit operational definition of Parsimony is:

$$\frac{\frac{\text{favorable}}{\text{possible}}}{\text{premises}}.$$

Precision is defined as a ratio of $\frac{\text{explained}}{\text{unexplained}}$ effects. Unexplained effects are considered to be error. Another way to state this definition is as a ratio of $\frac{\text{explicit effects}}{\text{implicit effects}}$. This suggestion explicates increased Precision. If explicit, or explained, effects are considered as favorable outcomes, and implicit effects, or error, are considered as unfavorable then precision operates as a probability of relative frequency. Relative frequency statements (odds) suggest that a person should never have more faith in an outcome than the evidence merits. In this case more precision is not necessarily better. Optimum precision is determined by how much faith is needed for any particular Conclusion.

Precision is applied to either a single equation or to a simultaneous set of equations. Operationally, precision is the ratio of explicit effects to error (implicit effects).

$$\frac{\text{effects}}{\text{error}} \quad \text{or} \quad \frac{\text{favorable}}{\text{unfavorable}}$$

More precision either requires adding additional terms to the equation or equations to the set (explicating a portion of the error term); or assuming a priori, zero coefficients among some of the dependent variables (assuming null relationships). This is pointed out in Blalock (1, p. 59-65) as part of the problem of identifying systems of equations. Morrison (16, p. 53) states that a system is inconsistent if it doesn't meet minimum precision. These two references indicate that a lower limit on precision is the limit needed to just identify a system or to form a consistent system.

Accuracy applies to the explanations of a theory. This dimension lies within the set of premises. Accuracy addresses the question: which set of premises is most accurate when several available sets appear equally plausible. Operationally this is the ratio of: $\frac{\text{statement}}{\text{all other statements}}$.

This is a case of the Probability of Confirmation. The conclusion does not follow necessarily, but only to a particular degree. This ratio is synonymous with plausibility.

Additional accuracy requires either Agreement or Differences as a strategic test. An example for the method of agreement is:

ABCD --> E ABFG --> E ACGK --> E	BCD do not occur in all 3 left hand terms. Therefore, BCD are not considered necessary to generate E. A, on the other hand, occurs in all 3 left hand terms. The conclusion is that A <u>may</u> be necessary.

An example of the method of differences is:

ABCD --> E $\bar{A}$BCD --> $\bar{E}$	Where A? generates E but not A ($\bar{A}$)? generates not E ($\bar{E}$). The conclusion is that A <u>may</u> be sufficient.

In either method, agreement or differences, the Probability of A being accurate in relation to other statements increases even though it has been proven to be neither necessary nor sufficient. This discussion provides both theoretical and operational definitions for the four borrowed dimensions.

Dimensions, as defined here, are continuous but contentless. Content is supplied by the Elements defined previously. Elements are categorical terms. Multiplying each Dimension with each Element (crossclassifying) yields potential Continuous Terms. These continuous terms specify potential theoretical and operational definitions.

The four dimensions (Scope, Parsimony, Precision, Accuracy) form one side of a two-dimensional array. The five borrowed elements (Resources, Structures, Integrations, Performances, Outputs) form the second side. The 4 X 5 array that results from the cross-classification provides 20 cells or boxes which make-up the initial solution grid or matrix. This strategy moves the inquiry from a descriptive definition to a prescritive framework. Each cell provides both dimensionality or variability and attributeness or content. Each cell then is a potential General Variable (a variable is considered to be general when it is free of places and times).

The lack of specified statements (connections, premises and equations) on Evaluation in general and Criteria of Criteria in specific, prevents further definition of Parsimony, Precision and Accuracy at this time.

Labels and Meanings have been tentatively established and these can be applied to establish verbal statements and premises. Before continuing to that stage however, a test of suggested labels and meanings was conducted. This was carried out to prevent continued efforts with faulty or incomplete Terms.

DATA

Test of Names and their Meanings

A literal definition (label) was inserted into each cell of the matrix. This provided both a theoretical definition (dimension by element) and a literal definition.

ELEMENTS

DIMENSIONS	INPUTS	STRUCTURES	INTEGRATIONS	PERFORMANCES	OUTPUTS
SCOPE	Utility	Generality	Applicability	Importance	Completeness
PARSIMONY	Explicitness	Simplicity	Matchability	Necessities	Consistency
PRECISION	Reliability	Univocality	Orthogonality	Objectivity	Confidence
ACCURACY	Truth	Validity	Soundness	Explanations	Power

This 4 x 5 matrix brings out nine primitive terms (5 elements and 4 dimensions). Since the matrix cross-classifies elements by dimensions the maximum possible number of derived terms is 20, assuming null interactions, which is assumed for the moment. (During the discussion at St. Louis it was suggested that we cross-classify elements by elements. This allows for many more primitive terms at a more specific level. This will be discussed in a later section of this report.)

A listing of definitions, both theoretical and literal, provides the following:

THEORETICAL DEFINITIONS		LITERAL DEFINITIONS
scope of inputs	=	utility
scope of structures	=	generality
scope of integrations	=	applicability
scope of performances	=	importance
scope of outputs	=	completeness
parsimony of inputs	=	explicitness
parsimony of structures	=	simplicity
parsimony of integrations	=	matchability
parsimony of performances	=	necessities
parsimony of outputs	=	consistency
precision of inputs	=	reliability
precision of structures	=	univocality
precision of integrations	=	orthogonality
precision of performances	=	objectivity
precision of outputs	=	confidence
accuracy of inputs	=	truth
accuracy of structures	=	validity
accuracy of integrations	=	soundness
accuracy of performances	=	explanations
accuracy of outputs	=	power

Generally, theoretical definitions are easier to understand and agree to than are labels or literal definitions. Each respondent prefers a particular label for each cell. Those suggested here are simply the label which this investigator believed captured the most meaning for him. As such they are only one set of many possible sets. This set of labels was the tentative set designed prior to the St. Louis discussions.

A completed definitional listing would of course include operational definitions or measures. Following the rule that two perspectives are sufficient to provide a cross check on each other greater efficiency is obtained by comparing each available pair of definitions rather than attempting to identify a third perspective for each of the Terms. It was with this in mind that the listing of theoretical definitions and literal definitions was used as initial input to the St. Louis discussions.

The elements of Resources, Structures, Integrations, Performances and Outputs are general and apply to any system. More specific elements for a system of Evaluation were borrowed from Morris (15, p. 466). This analogy followed the lead given by those who point out the root of the word evaluation is value. Morris lays out five philosophies of western man. The commonalities of these five are the existence of ontology, epistemology and axiology as elements. These elements were superimposed over the borrowed elements of the more general system to yield:

RESOURCES	STRUCTURES	INTEGRATIONS	PERFORMANCES	OUTPUTS
ontology	epistemology	axiology	???	???

The assumed sequence for social systems is that Resources are Structured and then Integrated. The assumed sequence for philosophies is that a particular view of reality (ontology) impacts upon the way knowledge is structured (epistemology). Epistemology in turn limits what is seen as good (ethics) and what is beautiful (aesthetics). Ethics and Aesthetics are considered here to be subcategories of Axiology. No philosophical elements were identified at this stage for either Performances or Outputs.

The matrix that results from cross-classifying these six elements with the four previously defined dimensions provides a prescriptive shopping list of 24(4 X 6) possible continuous terms as theoretical statements which may or may not apply to Criteria of Criteria. It is so far only a checklist -- each cell does not necessarily require an entry. As a check on their application, theoretical definitions are provided by combining each row entry with each column entry. These theoretical definitions are:

Scope of ontology
Scope of epistemology
Scope of ethics
Scope of aesthetics
Scope of performances
Scope of outputs

Precision of ontology
Precision of epistemology
Precision of ethics
Precision of aesthetics
Precision of performances
Precision of outputs

Parsimony of ontology
Parsimony of epistemology
Parsimony of ethics
Parsimony of aesthetics
Parsimony of performances
Parsimony of outputs

Accuracy of ontology
Accuracy of epistemology
Accuracy of ethics
Accuracy of aesthetics
Accuracy of performances
Accuracy of outputs

ERIC

The first test conducted at St. Louis was for literal definitions for each of the four suggested dimensions (primitives). Since dimensions are content-less, Criteria was considered as the system under discussion. Thus literal definitions were suggested for Scope of Criteria, etc. This was conducted as a presentation work session at the Adult Education Research Conference, St. Louis, April 16, 1975. Fifteen of the 32 persons present shared suggested labels. Past experience led to the writer's expectations of 4 labels from each respondent; 12 of the 15 who turned in the worksheets met or exceeded this expectation. However, 17 participants chose not to turn in their worksheets. The labels for each dimension were ordered according to level of generality and for consistency with accepted usage (19). In each case generality (the more the better) was sought. Labels were included when usage was consensual in the literature of theory development. Unacceptable labels we explained as exceptions and were not included in further analyses. These labels suggested are summarized as:

Scope of Criteria

size (C)
largeness (C)
quantity (C)
extent (2) (C)
specificity (C)

inclusiveness (4) (I)
necessary criteria (I)
part to whole (I)

cover 96% of situations (G)
spectrum (G)
theoretical breadth (G)
conceptual breadth (G)
conceptual dimension (G)
variance of thought or behavior (G)
comprehensive (G)
restrictiveness (G)

focus (B)
parameter (B)
boundary (B)
circumscribe (B)

complexity (?)

KEY

1. Number in parenthesis denotes number of participants listing that particular label, if it was more than one.

2. C = completeness
   I = inclusiveness
   G = generality
   B = boundary
   ? = non-consensual usage

Of the 25 labels suggested:

6 suggested completeness
8 suggested generality
6 suggested inclusion
4 suggested boundaries
1 (complexity) didn't suggest any aspect of scope because scope and complexity are considered by this writer to be independent

No clear cut distinctions appeared among completeness, generality and inclusiveness. Any of these 3 (20 responses) portray the idea of scope. Boundaries also apply to space rather than abstract relations (19).

Parsimony of Criteria

productivity (A)
adequacy (A)
scientific (A)
practical (A)
mathematical (A)
utilitarian aspect (A)
necessary and sufficient A and B (A)

common terminology (S)
simple to understand (S)
easy to explain (S)

limitation (?)
selection (?)
generalizability (2) (?)

economic (2) (E)
effective use of resources (E)
input/output (E)
product/time (E)
product/intent (E)
time (E)

KEY

A	=	Adequacy
S	=	Simplicity
E	=	Economy
?	=	seemed more in line with Scope than with Parsimony

Of the 22 labels suggested:

3 suggested simplicity (S)
8 suggested careful use of resources or economy (E)
7 suggested adequacy (A)
4 were questionable (?) and appeared to apply to scope rather than parsimony.

Simplicity appeared to suggest plainness or uncompounded state. Economy suggests careful and prudent use of resources. Adequacy on the other hand, suggests a sufficiency of the end product. This provides the often cited ratio of inputs/outputs.

Precision of Criteria

exactness (E)
pin pointing (E)
on target (E)
reassurable (E)
.05 (E)
sig confidence level (E)

KEY

E	=	Exactness
F	=	Formalness
?	=	questionable meaning

exemplifying (F)
usefulness (F)
directness (F)

can be 100 at 0 (?)
descriptive (?)

specificity (2) (?) Scope
congruence of data to criteria
validity > Accuracy
non-overlapping sets of criteria
x, x_2, x_3, x_4 > Parsimony?

Of the 17 labels suggested:

6 suggested exactness (E)
3 suggested formalness (F)
2 are questionable (?)

Specificity, mentioned twice, is more akin to scope as previously defined. Non-overlapping sets, mentioned twice, suggests a parsimony dimension. Validity and congruence suggests accuracy rather than precision. These conclusions point out the dual themes of exactness and formalness as tapping the domain of precision.

Accuracy of Criteria

rightness (C)
correctness (C)
correctness of fit (C)
congruent with reality (C)
validity (C)
measurement compared to
actual performance (C)

KEY

C	=	correctness
?	=	questionable meaning

appropriateness (?)
direction (?)
target (?)
cannot be interpreted without
data or a scale (?)

prediction (P)

applicability (S)
exclusiveness (S)

Of the 15 labels suggested:

8 suggested correctness (C)
4 were questionable (?) and not particularly meaningful

Applicability (S) and exclusiveness (S) suggest generality or scope. Predictive (P) is associated with precision rather than accuracy. These literal definitions suggest:

completeness generality and inclusion	as components of scope
simplicity adequacy and economy	as components of efficiency
exactness and formalness	as components of precision, and

correctness as the sole component of accuracy.

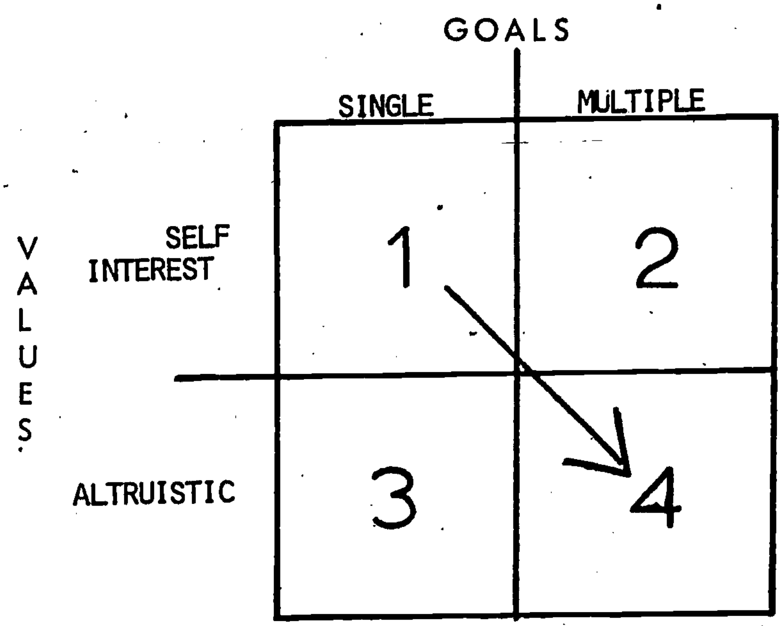
The discussion following the handout suggested three major shortcomings: First that the labels were "value-laden". This writer has no defense against such a suggestion. This objection points directly to an issue of inquiry. Is the goal to become value-free or is the goal to become objective? To some this is a trivial point because they assume it is impossible to be both value-laden and objective. These same people implicitly conclude the way to be objective is to become value-free. This is suggested as the logic behind the objection to specific labels because they are value-laden.

As the design matrix explicates, a goal of evaluation is objectivity (precision of performances). Nowhere to this point has value-freeness been suggested as a goal of Criteria of Criteria. Based on this and the resulting objection of value-laden labels the crux of the argument centers on whether or not it is possible to be both non-neutral (value-laden) and objective.

Following the lead suggested by Hage (7), this writer's answer is a resounding Yes--the classical literatures notwithstanding.

In general, it is assumed that people can either hold a single altruistic goal or hold multiple self-interest goals. As portrayed in a 2 x 2 array this suggests either cell 2 or cell 3.

ERIC



The possibility of being both non-neutral and objective lies along the main diagonal (cells 1 and 4). Is it possible to move from cell 1 to cell 4 as the arrow indicates? Specific values are non-variable concepts. They lead to stereotyped either-or debates, pitting "us" against "them".

To move from cell 1 (specific values and single goal) to cell 4 (general values and multiple goals) requires both a shift from categorical to continuous thinking, and a shift from single to multiple goals.

Both thinking in continua and thinking from multiple perspectives lead toward non-neutral objectivity because of the following:

1. Continuous concepts lead to adopting an evolutionary rather than a revolutionary perspective.

2. Continuous concepts force you to think in extremes--then it becomes impossible to be an extremist.

3. A shift to multiple variables is needed to capture all of reality.

4. A shift to a perspective of multiple values prevents ends from justifying the means.

5. Multiple values and multiple variables allow a move toward alternate ways of reaching the same end.

6. Multiple alternative ways and multiple ends approach objectivity without neutrality.

7. As you increase the number of variables or the number of alternative pathways the number of internal checks increase geometrically.

8. Thus it becomes harder to distort.

9. Internal consistency allows for non-neutral objectivity.

The design strategy suggested in this paper addresses the value-laden question head-on. A joint attack of variable concepts and multiple perspectives is suggested as a way of penetrating this paradox. On the other hand, this objection can certainly be expected since no measures were forthcoming. This is not a weakness of the approach but rather a weakness of its specific application here. Second, the question was raised as to cross-classifying structure and function with the elements of ontology, epistemology and axilogy. At first glance this suggestion was rejected (structure was based as epistemology) and function was part of the system, but had no completed cells. More thought however indicated the heuristic value of crossing the philosophical elements with the system's elements. This is given by the matrix.

	INPUTS	STRUCTURES	INTEGRATE	PERFORM	OUTPUTS
ONTOLOGY	ontological inputs	ontological structures	ontological integrations	ontological performances	ontological outputs
EPISTEM-OLOGY	epistemol inputs	epistemol structures	epistemol integrations	epistemol performances	epistemol outputs
ETHICS	ethical inputs	ethical structures	ethical integrations	ethical performances	ethical outputs
AESTHETICS	aesthetic inputs	aesthetic structures	aesthetic integrations	aesthetic performances	aesthetic outputs
PSYCHOL-OGY	psychological inputs	psychological structures	psychological integrations	psychological performances	psychological outputs

The fifth row (psychology) was added after further discussion about the matrix. This suggestion sprang from the inability to suggest adequacies of performance and the historical precedent of Psychology as a branch of Philosophy. As each cell is examined the question is asked:

What are (row), (column)? For example, what are Ontological Inputs (row 1, column 1)? Notice the cell entries are categorical (non-variable) terms. This is the case because this matrix is crossing elements with more elements which can only yield more categorical terms. These can be appropriately used for investigative units, elements of an investigative unit or modal cut points along a continuum (qualities). The main interest here is identifying elements for the investigative unit criteria for criteria sets. This matrix provides 25 possible elements. Before automatically using all 25 and re-applying the four previous criteria of scope, parsimony, accuracy and precision (which would yield 100 potential general variables) each element has to "make sense" or fit into a logically consistent system.

The strategic question is simply, which, if any, of the 25 possible elements are necessary for defining criteria for criteria sets as an investigative unit?

One approach is to explicate the definitional premise: Bench marks or comparative reference points to judge worth of a set of criteria?

The third criticism of this design strategy was that perhaps it really doesn't portray multiple perspectives at all--but rather says the same thing in three different ways. Since no operational definitions (measures) were developed this criticism became one of deciding if literal definitions (labels) were synonymous with theoretical definitions (meanings). The literature of theory development suggests this is not the case as do this investigator's experiences. As a result of the discussion in St. Louis it is suggested that the distinction was blurred in communicating the idea, but the idea of two distinct definitions is supported by the discussion. Such comments as "it is easy to see we all agree on the theoretical definitions but the labels mean specific things to each of us"; or "labels are too value-laden"; or "it seems to me you are merely playing with words", suggest the discussants sensed the differences between labels and meanings, but failed to move toward verbalizing the differences by concentrating either on labels or meanings. No single person appeared to attack or support both conjointly.

SUMMARY

As this discussion shaped out, there was a vague uneasy feeling that it was not achieving closure. That is true of course, but perhaps irrelevant. For if this is truly a dialectic creative-critical effort, this writer's experience suggests it requires a minimum of two rounds and usually three to achieve tentative closure. The discussion during St. Louis AERC completed the first round. Your written critique of the points discussed in this manuscript will complete the second.

Looking back it appears that the first question raised (how can you recognize a good criteria set?) still remains open. Tentative answers to remaining questions appear more explicit.

How can you develop a good criteria set? Applying techniques of continuous thinking and multiple perspectives has been successful for this investigator. To date these ideas hold the most promise of any for gaining scope, parsimony, precision and accuracy.

Where do good criteria sets come from? They come from explicit and logical techniques of analysis and synthesis. So far that is merely a personal opinion of this writer based on personal experiences. Following the systems approach there are assumed to be alternative ways to achieve any end. Finally, who can design a good criteria set? Anyone can! That includes you. That includes me. This follows, providing logical and explicit techniques can be communicated between people. Once techniques are explicated, and are logical, setting criteria can no longer be considered an art. Rather than subtle, sophisticated and implicit, the goal is convincing, simple and explicit.

> "Truth emerges more readily from error than from confusion."
> (Francis Bacon)

Your critique as cycle two of our communication should hopefully point to our errors.

REFERENCES

1. Blalock, Hubert M., Theory Construction from Verbal to Mathematical Formulations, Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1969.

2. Burris, Russell, "Human Learning", Chap. 4, Handbook of Industrial and Organizational Psychology, ed. Marvin Dennette, Chicago: Rand McNally, in press.

3. Dubin, Robert, Theory Building, New York: Free Press, 1969.

4. Economic Education Bulletin, "Plans for the Future", pp. 11-34, Vol. 8, No. 6, June 1973.

5. Gurland, Robert H., "Practical Reasoning", notes from Sunrise Semester, New York, N.Y.: New York University, Spring, 1974.

6. Hage, Jerald, Techniques and Problems of Theory Construction in Sociology, New York: John Wiley, 1972.

7. Hage, Jerald, Notes from discussions in Theory Building Seminar, Sociology 772, Madison: University of Wisconsin, 1972.

8. Hage, Jerald, and Gerald Marwell, "Toward the Development of an Empirically Based Theory of Role Relationships", Sociometry, pp. 200-212, June 31, 1968, Vol. 31.

9. Johnston, James T., et al., The IDEALS Concept Development Process for any Educational Situation, Madison: IDEAL SYSTEMS INC., undated.

10. Kreitlow, Burton W., Educating the Adult Educator: Part 2. Taxonomy of Needed Research, Theoretical Paper No. 13, Madison: Wisconsin Research and Development Center for Cognitive Learning, May, 1968.

11. Martin, James L., Before the Labels Fade: A Working Paper, Minneapolis: School of Public Health, University of Minnesota, May 1974.

12. Martin, James L., Open Learning Systems, Minneapolis: School of Public Health, University of Minnesota, November 1974.

13. Martin, James L., Developing a Perspective for Evaluating Open Learning Systems, Minneapolis: School of Public Health, University of Minnesota, August, 1975

14. Martin, James L., Strategies of Inquiry, Minneapolis: School of Public Health, University of Minnesota, in press.

15. Morris, Van Cleve, Philosophy and the American School, Boston: Houghton Mifflin Co., 1961.

16. Morrison, Donald F., Multivariate Statistical Methods, New York: McGraw Hill, 1967.

17. Nadler, Gerald, "A Universal Approach to Complex System Design", The Engineering Manager: Survival in the Seventies, 17th joint Engineering Management Conferences, Montreal, Canada, October 9-10, 1969, Montreal: Engineering Institute of Canada, pp. 121-139.

18. Nadler, Gerald, WORK DESIGN, A Systems Concept, Revised Edition, Homewood, Ill.: Richard D. Irwin, 1970.

19. Roget's International Thesaurus, ed. Lester V. Berry, 3rd ed., New York, 1962.

20. Stufflebeam, Daniel L., et al., Educational Evaluation and Decision Making, Itasca, Ill.: Peacock Publishers, Inc., 1971.

21. Worthen, Blaine R., "Competencies for Educational Research and Evaluation", Educational Researcher, Vol. IV, No. 1, January 1975, pp. 13-16.