

**MICHIGAN STATE**  
**U N I V E R S I T Y**

The Michigan State University Libraries are pleased to supply this material to you from our collection. The MSU library does not hold copyright of this material and cannot authorize any further reproduction.



**Interlibrary  
Services**

**MSU LIBRARIES**

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research". If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use", that user may be liable for copyright infringement.

If you have questions,  
please contact your library.

Michigan State University ILL (EEM)



ILLiad TN: 1273803

ILL Number: 205245567



Borrower: EXW

Lending String: \*EEM,GPM,TET

Patron:

Journal Title: Educational product report.

Volume: 2 Issue:

Month/Year: 1969

Pages: 36-38

Article Author: Scriven

Article Title: An Introduction to Meta-Evaluation

Imprint: [New York] : EPIE, [©1968]-©1972.

Notes: Billing Notes: FEIN# 38-600-7327

OCLC

Rec Date: 11/10/2020 11:32:59 AM

Call #: LB 1028 .A1 E2 v.2 Oct.-June  
1968/1969

Location: MSU REMOTE STORAGE B

AVAILABLE

ISSN: 0013-1865

Charge

Maxcost: 15.00IFM

Shipping Address:

Western Michigan University

RSC-Waldo Library

za015 \*\*\*\*\*

Melcat

Fax:

Ariel:

Odyssey: 206.107.44.66

Scan in ILLiad

B9-8-7-1

# **AN INTRODUCTION TO META-EVALUATION**

By Michael Scriven

Meta-evaluation is second-order evaluation, i.e., evaluation of evaluation. Theoretically, meta-evaluation involves the methodological assessment of the role of evaluation; practically, it is concerned with the evaluation of specific evaluative performances. Both facets will be dealt with here.

Evaluation is one of the most important, and from many theoretical as well as practical points of view, the most important function of science. The arguments for keeping science value free are in general extremely bad, but in the applied sciences these arguments have never even been significant, since it was always conceded that "instrumental evaluation" (i.e. evaluating alternative means of accomplishing given ends) was legitimate. Evaluation of educational products or processes is obviously a case of applied evaluation. Why, then, has there been much reluctance to undertake it in a straightforward and tough-minded way? Remember that this is a period in the history of the social sciences when tough-mindedness, in the form of empiricism, positivism and operationalism, has been regarded with much favor in the methodological sphere.

There are two explanations. The creditable one is that in the educational area achieving *general agreement* about the aims of education is not easy. The discreditable reason is known by the non-technical name of chicken-heartedness. A lot of people are threatened by evaluations of their products just as they are threatened by evaluation of their own performance, and they often take or threaten strong countermeasures, as was demonstrated by the much discussed battery additive case several years ago.

But the general and more creditable objection, often said to be "realistic," is equally indefensible. It is, of course, unbelievably hard to get people to agree about the "aims of education," and indeed it may be foolish even to attempt this. But that point is almost completely irrelevant to the one at issue. It is extremely difficult to get people to agree on a formulation of the goals of man's existence, or of morality, but this doesn't mean that we have difficulty in agreeing that kicking little children to death because they irritate one is morally indefensible.

An educational evaluation of "our schools" does indeed

---

*Michael Scriven is Professor of Philosophy at U.C. Berkeley and is on the board of directors of the Social Science Education Consortium. He has been a Fellow of the Center for Advanced Study in the Behavioral Sciences, a consultant for RAND on computer comprehension, and is on the board of consultants for the International Journal of Psychiatry.*

get into the area of debatable goals, though even then there is enough overlap between differing judgments of the goals of education to enable a very widely acceptable and strongly negative case to be made against the performance of the schools. But when we get down to specific issues, like the evaluation of overhead projectors, for example, or the evaluation of two programmed texts in multiplication for a specific age level and state curriculum, then the sting of this general difficulty is almost completely lost. We are, in fact, simply into the area of *Consumer Reports*, and anybody who thinks that what they're doing is logically or methodologically illicit needs his head examined. Of course, their evaluators sometimes make mistakes with respect to the criteria as well as with respect to the experimental design, but that happens to be the heritage of science and hardly an exclusive sign of sin.

It should be clearly recognized that attacking evaluative enterprises in the applied areas on the grounds that they are evaluative and hence simply matters of opinion or taste is not merely wrong-headed, but in many areas highly antisocial. The difficulties which EPIE had when it began are strictly comparable to those which beset Consumers Union in its early days, because of producers who did not wish to have their products evaluated. It is hoped that we have seen the last of that kind of outrage. For business enterprises, in a society which provides them with immense subsidies and support in the taxation and legal areas and which is committed explicitly to free speech and free enterprise with reasonable constraints, to attack a serious attempt at objective verification of their claims and publication of the result is to attack the basic axioms of free speech and fair play which support these enterprises.

In the educational field such attacks are particularly antisocial as most of the expenditure here comes from the public purse, and as attempts to evaluate these products are in the public interest. To put the whole matter in a different way, I think it should be continually stressed in the field of education, as in many other fields of research in the social studies, that any attempt to avoid evaluation should be viewed with suspicion, as probably a sign of methodological muddle-headedness or of a lack of social consciousness. The burden of proof for avoiding evaluation with all its dependent difficulties and risks must be upon those who wish to do so. The present debate about grading students' work on the campus and in the schools are typical examples of a combination of muddleheadedness and chicken-heartedness. It is perfectly clear that a great deal of the grading that goes on is superficial and unreliable and used as a substitute for pedagogically more satisfactory procedures. It is equally clear that the role of grading as (a) providing feedback to the

student on his progress towards the intended goals, (b) providing feedback to the instructor on his success in communicating learning, and (c) a basis for selective distribution of positions in specialist classes and types of reward or remedy is *absolutely indispensable* to the educational process in most areas and at most levels.

Evaluation does not begin at the point when the customer announces the values he wishes you to use as bench marks, except in a chronological sense. Methodologically and morally, evaluation absolutely requires that the evaluator investigate the justification for these "initial" criteria. Any evaluator who has substantial practical experience will know that frequently a project will be ruined by the discovery that the initially announced criteria turn out—when their consequences are realized—not to be the "true" criteria of the customer. The time to find this out, at least to the considerable extent that one can do this, is before investing heavily in the experimental design. And there are many ways to investigate both the phenomenological relative strength and the logical relative strength (i.e. defensibility) of the bench-mark criteria.

• • •

Turning to the practical applications of meta-evaluation, let us consider the specific assessment of "A System for Analyzing Social Science Curricula" by W. W. Stevens Jr. and Irving Morrisett, *The EPIE Forum* (Vol. I, Nos. 4 and 5).

The basic defects of the Curricula Analysis System are defects of clarity and presentation rather than defects of comprehensiveness. But these defects reflect a lack of clarity on the part of the constructors about the most important methodological distinctions to be made in evaluation, and unless they are corrected they will continue to result in deep-seated reactions by the consumers, who call the system "unwieldly in its present form" ("two-thirds of one clinic using it," p. 15). This system is a pretty good example of exactly what a programmed text writer learns to avoid very quickly. It should have been presented in alternative simple forms, e.g. with the major sections in various order. Most of the complexities have probably developed because the constructors of the system have followed the practice of gradually developing their own jargon and distinction. This is pedagogically bad since there are no reasons to suppose that these distinctions or this jargon are natural for the ordinary language user; consequently he has to acquire this new, and in my view, perverse language before he can use the system. Common though this situation is in the educational field, it is long overdue for radical reform, especially as the reform is not particularly difficult. Without personally doing the field-testing, I can only speculate about alternative terminology, etc. But as a critique of an

evaluative procedure, the main point I want to make is the one above—that field testing of preliminary, simplified forms are absolutely essential if one is not to develop an excessively metaphysical system.

The first question to ask is, "What does the consumer most want to know?" There is clear evidence on this presented in the article (p. 15) and the evidence makes it obvious that knowledge of rationale and objectives is the primary concern of the consumer. Therefore it should be the first section and not the second. It would probably be a good idea if a very brief introductory summary occupying not more than a paragraph and containing in the last sentence a summary of the overall evaluation was to lead off the entire analysis. We could then go to the more detailed subheadings under "Rationale and Objectives."

Having transposed the first two sections we now find that the original opening section, which is called "Descriptive Characteristics" is obviously mistitled. It is clearly a description of a curriculum to give its objectives and rationale. Indeed "Subject Area and Content" is a subheading under the original "Descriptive Characteristics." We may dispense with this in the revised order, and retitle the original Section 1 "Product Characteristics." We should then retitle Section 3 ("Antecedent Conditions"), "User Characteristics." Section 4 "Content," should be called "Details of Content," since a very good idea of the content was given by the "Objectives," and 1.8 "Subject Area and Content," subsection. And then we can wind up by rechristening Section 6 ("Overall Judgments"), "Evaluative Data," since it is not just judgments but data.

Under the new sequencing and titling, we have a much more plausible developmental sequence, but a number of the subsections need to be shifted around. Without going into immense detail, let me mention one or two examples where the present location illustrates a methodological misconception—in my view. We notice, for example, that subsection 1.7 "Performance Data Availability" is put in under "Descriptive Characteristics." But it's clear that this is simply evaluative data, and ought to go into that section. Of course it's also descriptive, because there is no significant distinction between description and evaluation in general. But in a particular context we make that distinction all the time, and the location for this section is clearly in the section stating the basis for the overall evaluation.

The use of the term "articulation" as the heading for 3.5 is an example of typical jargonistic performance. This is not the proper use of the term, and it is totally unnecessary to introduce a technical term at all at this point, since the heading "Relation to Other Curricula" completely covers the specific descriptions.

The real problem is in the detailed description of content for Section 4, where we find a distinction between the two subheadings one being called "Cognitive Structure" and the other being "Affective Content." As we read the detailed description we discover that the authors are committed to the view that values are not cognitive. This is a striking example of the contamination of evaluative schema by a dubious set of methodological prejudices. At an even simpler level, it is entirely inappropriate to identify affective content with values. Affect is feeling; values are theoretical constructs extracted from and causing committed behavior.

Naturally feelings are often associated with values, but they are simply different things, indeed different categories of things. One may value honesty or truth, but display no perceptible affect in discussing them.

Another distinction should probably be made that is not explicit in the system at the moment. Probably it would be desirable to rechristen the section on product characteristics "Product Characterization" or "Claimed Characteristics," and make clear that this description is *the one provided by the author or publisher*. Thus for a validation of the success claims built into the description, the system user should turn to the evaluation section, or simply face the fact that a large part of the descriptions of style, time needed for use of the materials, etc., which are supplied by the producers are entirely suspect and have no place in a factual description of the materials, except as a factual description of the originators' intentions until they have been verified experimentally.

In short, either build the evaluation back into the description where you will otherwise be faced with repeating a plug, or make clear that you *are* only repeating a plug in that section of the analysis. Many of my earlier criticisms of the long form, communicated to the authors, were on this point and have been incorporated in the modified short form published in the *EPIE Forum*, but there is still an absolutely pervasive sense of confusion about evaluation and description. Take, for example, this description under 2.2 ("General Objectives") "What are the generalized student outcomes that can be expected from the use of these materials?" Obviously, this calls for an evaluation and a crucial one. It *must* be made clear whether this is what the producer *says*, or whether it is, in fact, part of the evaluation tucked back into the rationale sections.

The commitment to the Bloom and Krathwohl classifications is far too pervasive, especially in the "long form." These taxonomies are seriously defective, and much of the critical work since their original publication has documented this. But more seriously, they are also

excessively complicated and jargon-ridden. So it seems unfortunate to pull in jargon on top of jargon, or—speaking practically—to require learning these classifications in order to use the Curricula Analysis System. In particular the "Analysis/Synthesis" distinction is really pretty shaky and not particularly important. Of course, no system for analysis can avoid using somebody's taxonomy, but the message is to use the *minimum* taxonomy i.e. make only those distinctions which are absolutely clear and indisputable *and* important.

In general, then, this analysis system, truly valuable though the idea is, needs a great deal of improvement. It is too committed to methodological or psychological doctrines of excessive dubiety, and above all it is *excessively* complicated, repetitious, and imprecisely described.