DOCUMENT RESUME

ED 114 260 95 SE 018 801

AUTHOR Scriven, Michael
TITLE Evaluation in Science Education.
INSTITUTION ERIC Information Analysis Center for Science, Mathematics, and Environmental Education, Columbus, Ohio.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Mar 75
NOTE 14p.; Occasional Paper Series

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS Curriculum; *Curriculum Evaluation; Educational Research; *Evaluation; Evaluation Methods; Evaluation Needs; *Science Education; Scientific Attitudes; *Scientific Methodology; *Values

ABSTRACT

This report discusses the importance of five types of evaluation in science and science education. First, evaluation must be recognized as a key process within science, both pure and applied, and must be taught as an integral part of science education. Second, the applications of science must be evaluated not only as a social responsibility and a pragmatic way to select researchable areas, but also as a scientific matter in itself, and thus a proper component of science education. A third type of evaluation in science education is the analytical approach to values-laden issues in a way that identifies the values' assumptions and proceeds to investigate their evidential status. A fourth important evaluation in science education is the evaluation of science materials; this is similar in some respects to the evaluation of applied science. The fifth way that evaluation enters into science education is in the attempt to list and indoctrinate the student with the supposed values of the scientific approach, such as truth, tentativeness, and criticality. Also discussed is evaluation as a form of knowledge and a form of education, as well as the types of evaluation content and attitudes specifically related to science education curricula. (Author/MLH)

SCIENCE EDUCATION INFORMATION
REPORTS

OCCASIONAL PAPER SERIES

EVALUATION IN SCIENCE EDUCATION

By

Michael Scriven
University of California
Berkeley, California 94720

ERIC Information Analysis Center
for Science, Mathematics, and Environmental Education
The Ohio State University
400 Lincoln Tower
Columbus, Ohio 43210

March, 1975

CONTENTS

SCIENCE, MATHEMATICS, AND ENVIRONMENTAL EDUCATION INFORMATION REPORTS

The Science, Mathematics, and Environmental Education Information Reports are being developed to disseminate information concerning documents analyzed at the ERIC Science, Mathematics, and Environmental Education Information Analysis Center. The reports include four types of publications. Special Bibliographies are developed to announce availability of documents in selected interest areas. These bibliographies will list most significant documents that have been published in the interest area. Guides to Resource Literature for Science, Mathematics, and Environmental Education Teachers are bibliographies that identify references for the professional growth of teachers at all levels of science, mathematics, and environmental education. Research Reviews are issued to analyze and synthesize research related to science, mathematics, and environmental education over a period of several years. The Occasional Paper Series is designed to present research reviews and discussions related to specific educational topics.

The Science, Mathematics, and Environmental Education Information Reports will be announced as they become available.

Occasional Paper Series - Science

The Occasional Paper Series (Science) is designed to present reviews of literature or discussions related to specific topics or educational programs related to the teaching and learning of science. We hope these papers will provide ideas for implementing research, suggestions for areas that are in need of research, and suggestions for research design.

The ERIC Science, Mathematics, and Environmental Education Information Analysis Center has cooperated with the National Association for Research in Science Teaching in sponsoring this paper as a General Session presentation at the 48th Annual Meeting in Los Angeles, California, March 17-19, 1975.

This paper will serve as the basis for a journal article to be published in the near future.

Stanley L. Helgeson
and
Patricia E. Blosser
Editors

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to the National Association for Research in Science Teaching for critical review and determination of professional competence. This publication has met such standards. Points of view or opinion, however, do not necessarily represent the official view or opinions of either the National Association for Research in Science Teaching or the National Institute of Education.

EVALUATION IN SCIENCE EDUCATION

Michael Scriven
University of California
Berkeley, California 94720

Introduction

There are five crucial ways in which evaluation enters into science education, and none of them have received much discussion, so far as I know. There is a fifth way which does receive perfunctory discussion, but it is not so crucial. I want to turn the orientation of science education on its head with regard to evaluation, by arguing that nothing else is as important in science or in science education. (Compare this with the AAAS analysis of scientific method as a basis for science education, or with SCIS, etc.)

Outline

The first of these ways is through content. That is, evaluation must be recognized as a key process within science, both pure and applied. Hence, for adequate science education, it must be conceptualized, and it must be taught.

The second way is through the evaluation of applications of science. We have recently become well aware of the necessity for social responsibility on the part of scientists in selecting or continuing in particular areas of research. I am arguing for something more than that, namely, for treating the evaluation of applications as itself a scientific matter, and hence of course a proper component of science education (and probably of education for non-scientists).

The third way is through treating what has come to be called values education as part of science education. Values education, as I here conceive it, is the analytical approach to value-laden issues, especially moral issues, in a way that stresses the identification of the values assumptions--the underlying principles and attitudes--and proceeds to investigate their evidential status. These values may be prudential, scientific, or moral.

The fourth way is through the evaluation of science education materials. This can be seen as a particular case of evaluation in applied science, but its self-referent aspect and the especially complex process involved entitle it to special consideration.

The fifth way consists in the attempt to list and indoctrinate the student with the supposed values of the scientific approach, e.g., truth, tentativeness, criticality. This approach has been flawed by bad conceptual analysis (in that the proposed values have been very vague and incomplete, incorrect, or too narrow), and by the tremendous limitations on generalizability to specific applications in other areas inherent in the learning process with vague or narrow concepts. In short, by weak analysis and poor teaching methods.

This is not an artifically inflated list. For example, evaluation plays a crucial role in the administration of science, which does not really fit neatly into any of the above categories. But let's call it a case of applied science, since there is at least the term "management science," and note its particular importance as another self-referent discipline.

Evaluations as Knowledge

The first obstacle to the plan to get evaluation into the science curriculum is the fact that many scientists and science teachers think that judgments of value, worth or merit simply are not scientific expressions at all, that they should be seen as more akin to expressions of pleasure or autobiographical remarks. This on-the-face-of-it peculiar idea did not spring instinctively to their minds. It came from philosophers. An adequate treatment of the pernicious doctrine of value-free science is beyond our scope here, and I have covered it at great length elsewhere in an article called, "The Exact Place of Value Judgements in Science." What I shall do instead is to appeal to two authorities on the matter. First, your common-sense. Second, in case this is too near to Sunday for your common-sense to quite swing the deal, the Bible, which has some most interesting (may I say) revelations on this topic.

The common-sense of the matter goes like this. I assert that there could be no such thing as science, to be done or to be taught, if it were not for the fact that science was an evaluative activity. For science can only be distinguished from pseudo-science in one way: namely, by distinguishing between good evidence and bad, between good explanations and bad, between good experimental designs and bad ones, between good instruments and bad ones, between good arguments or inferences and bad ones. These are the value judgments on which the very possibility of identifying science rests. The teaching of science is very largely, even now, and implicitly, teaching the evaluative skills required to make these discriminations. The very words "science" and "scientific" are thus evaluative, meaning part of the body of well supported claims, as opposed to badly supported ones.

From long experience, I know that some of you will be saying to yourselves: but these aren't really value judgments, these are claims that can easily be translated into objective language. Absolutely correct! But that happens to be the very nature of value judgments. The catch is that no translation can be given that is independent of context. There is no translation of "good" that works in every context - but it's easy to give a pretty workable translation of "good top-loading balance for an undergrad quantitative analysis lab." Now, of course, as with any other crucial term in science, such as "true" or "explanation," there are times when the term is used and cannot be given adequate support. Thus, when people talk about good wines and good paintings, then --apart from the fact that they no doubt imply that the wine hasn't turned to vinegar and the painting isn't crumbling into dust--it's simply not possible to translate the claim into objectively testable components. This is no more a sign that value judgments are intrinsically subjective than the telling of a lie is a sign that the same is the case with truth. Despite the possibility of translation within a

given context, translation into descriptive language, evaluation is not the same as description. An evaluation is a kind of description, but evaluation is a special kind of description and most description isn't evaluative. Broadly speaking, evaluation is the pay-off (or a major pay-off) of description--it is the synthesis of descriptive information and criterial information according to very strict and difficult rules. Another pay-off of description is explanation--typically, it requires some description in terms of generalizations, sometimes some particular facts, and always some assessment of certain context-given variables such as levels of prior understanding. Out of these ingredients we synthesize scientific explanation, the vehicle of the scientific enterprise. In just the same way, we learn how to synthesize facts about different explanations with the criteria which explanations are supposed to meet, in order to yield estimates of their relative merit, in short, value judgments about explanations. There's nothing simpler about evaluation than about explanation, there's nothing less scientific, and there's nothing less important. Evaluation is an absolutely crucial part of science, just as crucial as observation and more crucial than explanation, generalization, or prediction, since we can have a body of important scientific data without propositions, generalizations, or explanations --but we cannot have it without evaluation, in this case evaluation of the quality of the data, via evaluation of reliability of the observers, the instruments used, and the combinatorial procedures.

When you know that refractors are better telescopes--ceteris paribus --than reflectors for planetary observations, you know something most important, something which took many years to establish, something which can be supported and explained just like any other scientific claim in astronomy. Of course, evaluations are knowledge, and of course they are important knowledge. How, then, does the entire concept of evaluation escape mention in most of the approaches to science education with which I am familiar and avoid serious treatment in all of them? The consequences of this omission and what it symptomizes are horrifying, for it is nothing less than this misrepresentation of science that has led to the excuse that science is not concerned with making value judgments, than it is concerned with description and not recommendation, with facts, not values, with the "is" and not the "ought"--the whole mythology of purity that has led to the misemployment of scientific research, its too-frequent trivialization and its too-common failure to be sufficiently concerned with issues that need attention and which could benefit our brothers and sisters who pay the bills and need the help. I'll come back to that charge and fill in some steps in the argument in a moment. But first, I want to take a few moments for a Bible lesson, which you may regard as entertainment or inspiration as seems appropriate. The theme is the knowledge of good and evil, the kind of knowledge that one can obtain scientifically from evaluation.

The Origins of Knowledge About Value

Here we are in Genesis 2:16 ff. --God is showing Adam around the Garden of Eden: "And the Lord God commanded the man, saying, Of every tree of the garden thous mayest freely eat: but of the tree of knowledge of good and evil, thou shalt not eat of it: for in the day that thou eatest thereof thou shall surely die."

God now creates Eve, who gets into a conversation with the serpent, who is described as being "more subtil than any beast of the field which the Lord God had made." Eve mentions the ban on eating the fruit of the values tree and God's warning of fatal effects, and the serpent says (Genesis 3:4-5): "...Ye shall not surely die: for God doth know that in the day ye eat thereof, then your eyes shall be opened, and ye shall be as gods, knowing good and evil."

So Eve bites into the apple and gives Adam some, and their consciousness is raised mightily, to the point where they intuit that nakedness is evil and don some figleaf aprons. God gets very upset when he finds out about this, and puts it to Adam who blames Eve who blames it on the serpent. God then punishes them all in various ways and in justifying the expulsion from Eden, God says (Genesis 3:22-23): "...Behold, the man is become as one of us, to know good and evil; and now, lest he put forth his hand and take also of the tree of life, and eat, and live forever: therefore the Lord God sent him forth from the garden of Eden.

Now, of course, the main point of this excursion is to stress the antiquity of the notion that the most potent knowledge is knowledge of good and evil. Since we <u>have</u> inherited this from Adam, however, we might take a moment to apply it to this episode. Nobody comes out of this very well if we look at their scorecards.

The serpent comes out best. He was, as it turned out, a little subversive, but in a charitable cause--he corrected an error; perhaps he can be seen as the first scientist. He made no recommendations to act against God's command, he merely stated what God later concedes to be the truth, namely that the apple was not fatal and that it would make Adam and Eve like gods in an important respect.

Eve comes off next best. She eats the apple on her own initiative, which seemed like a good idea at the time. But then she finks on the serpent.

Adam directly disobeys an order, not on the basis of discussion with an independent authority, and he starts the whole finking routine.

But, of course, God does worst. First, he lies about the apple's toxicity. Maybe that was in a good cause, but it seems like bad management to create a smart serpent who's going to spill the beans, or for that matter, a woman who would believe a serpent before God. If it comes to that, why leave the tree around at all? Was it a gigantic ethnics exam? If so, God flunks as an item-writer, since in that situation Adam and Eve acted pretty sensibly; even obeying the orders of someone that gives you a garden isn't quite as attractive as becoming a god yourself, and of course they had no knowledge of good and evil <u>before</u> they ate the apple, so they can't be blamed for <u>moral</u> error. Consequently, He was extremely unjust in His treatment of the serpent, who only spoke the truth; the woman, whom He punished more severely than the man; and the man. Moreover, He seems rather jealous since He kicks them out of Eden to prevent them becoming immortal and hence true gods; since there appear to be a number of gods already (God uses the plural), why not two more? Then there is the rather serious point that God's conception of good and evil (which is presumably what we get from

eating from His tree) seems a bit strange. The effect of acquiring a moral sense is that Adam and Eve get uptight about nakedness and start blaming others for actions that were clearly their own responsibility. Since God made them naked in the first place, it can't really be an intrinsically wicked state, so they're wrong about that; and the finking move is clearly immoral on most Old Testament rules. So it looks as if God is pretty confused about good and evil, Himself, quite apart from His very weird idea of justice, illustrated in the punishments He hands out. He's not even in very good shape about the other flora in Eden. Adam didn't get to the tree of life, in fact he gets thrown out just to keep him from it, but far from dying on the day that he ate the apple, he lasted nine hundred and thirty years (Genesis 5:5), which might be as near to "forever" as one would care to get. Nobody mentions Eve after her giving birth to Seth at age one hundred and thirty; in fact, nobody mentions any women in listing the generations that follow. It is probably not totally surprising that a number of women's groups are demonstrating against the Bible in churches this month.

Well, that's the problem with knowledge of good and evil--it's dangerous stuff. We get it from God, according to Genesis, and here we are turning it against Him. Or perhaps we get it from science--and before long we are using it to criticize science. Those of us who are professional evaluators are all too well aware of the contemporary dimensions of that danger, of the threats, the hostility, the paranoia, the blind resistance that one encounters in the field. To some extent, this is understandable, but to a large extent it is a result of either ignorance or impropriety. It is, in short, the reaction of those never educated to believe that knowledge of good and evil is legitimate, let alone valuable. It takes a real effort to develop a culture to the point where Consumers Union was blacklisted for more than a decade, unable to advertise in any newspaper (and I think any magazine) because it was felt by other advertisers--who convinced the publishers--that publishing evaluations of consumer goods was an attack on free enterprise and certainly on them. The first months of the EPIE Reports--the Educational Products Information Exchange, an effort to produce a kind of Consumers Reports in the educational field--were bedeviled by threatened lawsuits by manufacturers of A.V. equipment which was unfavorably reviewed. Not because the criticisms were unjustified, or not in the public interest, but just because they were criticisms. And this is just a few years ago. There are a few--but a few--signs of change. At the level of science education materials, whose development has been funded by the big agencies, the story is the same--surly resistance to evaluation, attempts to subvert it or co-opt it or censor it. This is still happening today, though the agencies themselves are beginning to get interested in evaluation. But the reason for this has been Congressional pressure, not scientific principles. Now where in the educational curriculum should there have been prophylactic treatment that prevented this self-impoverishing and socially harmful reaction by scientists as well as business people? There is no doubt that science and social studies, but most especially science, is the place. For science has the precision, the prestige, and the practice at evaluation, although of course it has been closet evaluation.

Evaluation as Education

I hope we have established the legitimacy of evaluative knowledge. Let us now review once more the various ways in which evaluation enters --or should enter--into science education, before we begin to look at specific curricular possibilities.

The first way is the one we have been discussing as our route to legitimation--evaluation is part of the practice of science, whether pure or applied, and as such, adequate science education must recognize it and include it.

The second way is specific to applied science, and to our particular kind of applied science, namely, science brought to bear on externally generated problems. This is often, but not always, called technology or engineering, but the new subjects of space science and computer science illustrate the fact that more than "mere" engineering is involved. The function of evaluation here is absolutely crucial, and consists particularly in the process of deciding whether a solution proposed by the scientist actually solves the problem posed by the external client. In simple cases this raises no need for a special discipline; for example, it's easy enough for the scientist herself or himself to decide whether the problem of predicting major earthquakes has been solved. But the situation changes radically when we start talking about the field trials of a naval weapons suite or an interactive smart terminal in a CAI system or an FM multiplex receiver for relay use. Here we need experts, evaluation experts. Sometimes these are quality control or product or procurement engineers who became good at evaluating specific products from considerable experience. But sometimes there emerge evaluation specialists with a considerable range of expertise--for example, the road testers on the staff of automobile magazines, the Hirsch-Houk hi-fi evaluation group, the FDA drug evaluators, Consumers Union, some computer consultants, and the educational evaluators. But evaluation is a crucial part of this kind of applied research, not only in the terminal role we have just been describing, but also in the developmental process. Science education, it seems to me, should spend more time than it usually does on the R&D process itself as an object of study, and that process of course involves the (formative) evaluation of each successive version of the product. The same process occurs within pure science, as a theorist refines a hypothesis, but in applied science the evaluation process is much more distinct from the scientific thought process and often becomes the job of an independent consultant, just because the scientist realizes the risk of bias in judging his or her own product. There are nice opportunities for role-playing exercises here, of a kind not too common in science education, but--I think--representing a valuable addition to what is provided. The roles of the scientist, of the external evaluator, and of the client are trickily interactive and form a microcosm of the general problem of the social accountability of science. And that problem represents the overlap between the second and third areas in which evaluation enters science education. For the early development of pesticides or nuclear reactors represent examples of good applied science but poor or underemployed evaluation. The client did not specify the requirements for the scientist with due regard to side effects and public costs, something a more careful evaluation would--and later did--pick up. (This illustrates

the utility of the evaluator in the planning phase, which one might regard as the limit case of formative evaluation.) The issue of the public interest in applied science is an evaluation issue.

But it also concerns the third area, namely the idea of values education itself. When we start moving into societal effects, we are moving into moral considerations. Now it might be the case that despite the fact that science intrinsically involves making value judgments, it does not involve making moral judgments. Any hope for that position fails when we start looking at applied sciences such as medicine and education, where the very definitions of the crucial concepts such as health, improvement, or educational progress, must involve a moral element. It would take us too far afield to explore the ramifications of this point, and the various responses to it that are possible. One might say, for example, that the applied scientist should accept the values of the society in which (s)he finds himself or herself. That position is not so popular today as it was a decade ago, but--popular or unpopular--it must be examined carefully against the alternative, which is to treat ethics itself as wholly or partly a matter for scientific analysis. I do not believe one can avoid taking at least the middle positions here, i.e., the position that ethics is in part a scientific matter. It is hard to act as if decisions about the death penalty, for example, are not in part dependent on the statistics about the deterrent effect of this punishment and alternatives to it. The real question is how large a part of ethics is amenable to scientific analysis. I think the extensions of game theory, welfare economics, and latent function analysis in sociology have now put us in a position to assert that a very substantial part of ethics, as of jurisprudence, can be approached as a scientific subject. But the size of the slice of the pie is not crucial. The simple fact is that a substantial part of what has been called values education recently, including most obviously the values-clarification component, must be regarded as a legitimate component of social science education and hence of science education in the broad sense. There are many people in science education who would love to see the subject kept clear of involvement with ethics--perhaps just by giving it to the NCSS people. There are others who want to go in the opposite direction, and I'm delighted to see you have a symposium on this issue scheduled for Tuesday afternoon. One way or another, however, I think we're going to see more and more involvement of science education in training students in the real analysis of social issues and not just as a handmaiden to the moralists; as an active collaborator in determining the best moral position as well as the best means to the moral end.

The fourth reason why science education can't remain aloof to evaluation arises from the steadily increasing pressure for more serious evaluation of science education itself. Now one can treat this as a kind of political imposition from the funding agencies--and it has often been treated in this way. But one can also see it as the demand that the physician treat himself, that science education should act like it is a scientific enterprise itself, that it use the R&D process on itself. This seems to me both logically and ethically appropriate. From this perspective there springs another great opportunity for science education --the opportunity to convert this necessity for evaluating itself into

an opportunity to use that process as an educational exercise itself. I'll elaborate on that in a moment.

The final intrusion of evaluation is one that I consider partly trivial and partly dangerous; it is the attempt to identify and perhaps indoctrinate students with the alleged values of the scientific enterprise itself--the search for truth, etc. If these are proposed as the dominant motives of scientists, then the proposal is naive, ill-founded, and objectionably self-congratulatory. Recent events at Rockefeller and Harvard are not the important data here--more can be read from the studies of contemporary science such as the recent history of the successful quest for a Nobel Prize in biology (The Double Helix). The real motivations of scientists are about as various as those of businesspeople or bureaucrats. If, on the other hand, these values are put forward as ideals, they have as little claim to an effective place in science education as have moral ideals in moral education--that is, they are totally inefficacious and only known to increase guilt, as likely to produce loyalty as saluting the American flag every morning. And to the small extent they are successful, the loyalty is likely to be a blind loyalty, no virtue at all.

Evaluation Content and Attitudes in the Science Education Curriculum

What do all these reflections translate into, as far as a science curriculum is concerned? I could hardly resist the temptation to say something on this issue, even if I didn't believe that a theory without applications has little chance of making converts, given the nature of this evening's program, where the luminaries of this organization will report on "Priorities for Research in Science Education," the results of the NARST-NIE Commission. I notice that two of the participants also have close ties to the NSF, so we may indeed get an authoritative view from those agencies. Let us sincerely hope that they will have a high priority for evaluation research on science education, and, if research on curriculum is part of their agenda, that they will have something to say about the possibility of evaluation as part of the curricula. For, despite the workshop on evaluation, aimed at teachers, I can't find anything else in the program that looks like a plug for the "evaluation curriculum." Dave Berliner of course could bring it in under a title like "Studying and Defining Effective Science Teaching"--but then he could bring almost anything in under that topic, which is probably part of his wisdom in choosing it.

Where would I bring it into the curriculum? In a dozen ways--it is compatible with most approaches--it can be part of the supplementary packages for most courses. But let me suggest that it can also be a kingpin in the structure. The phrases that SCIS, for example, use to describe their emphases are "exploration, invention, and discovery." I'd like to see "evaluation" in every short list of stressed concepts. I'd like to see a whole series of modules that focus on it specifically. Of these, a substantial number would focus on evaluation of non-educational entities--but a number would focus on science education itself.

I find that even very young people are pretty interested in certain types of evaluation. It's as good an entry point to their attention as their interests in watching animals or making things--or, to be more sensible about the evidence, it seems to be as good a start for some kids. They may be interested in evaluating baseball players or places to live or careers or people or games. As they get older, they get interested in evaluating cars and guns and garden fertilizers and consumer goods and food fads. In each of these cases, there's an important part of the evaluation that involves straight scientific research. But there's more than this; all the skills of experimental design to avoid or correct for bias, the notion of double-blindness and sample size and randomness, and special tests--all of them come in naturally and it is an easy transition to the valuation of scientific instruments via, e.g., thermometers with their hybrid role, partly in everyday life and partly in science. Then we can begin on the evaluation of scientific theories, scientists, research projects, and achievements.

And the step from that to the evaluation of texts, course materials and labs, science teachers and science students is a natural one. It would not be a bad thing for science education to raise the consciousness of students to the evaluation of these components of the school, to make this direct contribution to the life of the student outside the science classroom. Skill in the scientific evaluation of consumer goods is another kind of contribution that has both cash value and social value to a degree that much science education fails to provide. The nice thing about this approach is that it is consistent with the most serious and scientifically substantial research, as well as science in the public interest. The time has come to convert a political reform movement that has moved from consumerism to public interest legislation and accountability, into a change in the content of the curriculum--not somewhere else, but right here in science education.