Labs

Reviewing Reviews

Don Igelsrud
Department Editor

Previously I indicated that I hoped to develop a dialogue about laboratory teaching and the practical problems of biology teaching. I suggested this was necessary because there is so much material available for teaching that no one can try everything on his or her own. I ended by asking you to tell me what you thought was "best."

As I looked back on my experience as a biology teacher, to try and answer the questions I asked you, I had to review a wide variety of experiences. It was hard to choose a single item as best because there were many good things, some I particularly liked, and some students enjoyed more than I did because it was new to them. However, after some thought I came up with my answers. If you haven't responded to the questionnaire yet, please do!

Here are my choices:

Best Lab Experiment or Activity

In an earlier column, I described a lab exercise where students were given a living, young rat to study. They looked at behavior and adaptation and then studied the major stages of anesthesia as they prepared the animal for dissection. I considered it a best exercise because, when it is properly and humanely taught, students learn a great deal about the qualities of a living organism. I can think of many excellent exercises which illustrate specific concepts well, but I have chosen this activity because of the large amount of high quality experience it conveys.

Worst Lab Experiment or Activity

The dissection of a preserved starfish. The small amount of information conveyed compared to the amazing qualities of the living animal make this a terrible waste of an animal's life and student's time. There are many other animals I could have chosen which are equally bad.

Best or Most Useful Organism

The frog. This animal is easy to maintain and useful for a wide variety of fundamental experiments, many of which do no harm to the animal, e.g. color change, capillary circulation and feeding behavior. Physarum would be my next choice.

Best Film or Supplementary Material

This is a difficult category for me, because my answer could be viewed as biased and self serving. The "Principles of Biology" videodiscs would be my choice for this category and for the next one, which is most useful item. Clips from the best films I know of are on the discs, making them useful nearly every day one teaches.

My choice for best film is "Work of the Kidneys" from Encyclopaedia Britannica because it explains their function using living material and exceptional microphotography. The explanation is good and because dialysis is used to introduce the kidney's function, the whole discussion is placed in a relevant and human context.

Most Useful Item I Purchased for Teaching

The Balston air filter I bought to filter and control the pressure of compressed air coming to my lab. After it was installed, I had no trouble with aquaria because air pumps did not work properly. My next choice would be the Autolet "painless" lancet system for obtaining blood. In an almost completely painless way it allows one to produce a puncture which bleeds enough to get a good sample, avoiding all the problems students have using only a lancet.

Reviews are used by society to evaluate new items in the marketplace. Unfortunately, there seems to be more interest in producing more material to read than in trying to help consumers make good decisions about what to use. I would like to spend the rest of this column discussing those problems with the hope of bringing some light to them, and I would like to make some suggestions which I hope will make reviews more useful.

I do not wish to discourage publishers of reviews. It is a tremendous task to find qualified reviewers, to identify and send out materials, and to organize and edit the results for publication. What I would like to see, however, is more effort to develop standards so the reviews are more meaningful. Here is a simple example of the problem, one which I have chosen because many people will be familiar with it.

Milk Chocolate

Consumer Reports recently evaluated milk chocolate. It hired an expert who produced an evaluation, a review. The highest rated milk chocolate was an American product, Ghirardelli, which is made in San Francisco and is well known. The lowest rated product was Cote d'Or, which is made in Belgium. Because I had the pleasure of discovering Cote d'Or this past summer at the Belgium exhibit at Expo '86 and because I found my way to Ghirardelli Square for the first time the previous spring, I have evaluated these two chocolates as well as a variety of others over the past year. Consequently, I feel I can comment fairly on the *Consumer Reports* review.

One could argue that this is the same *Consumer Reports* that rated New York City tap water as the best mineral water, well above Perrier, and that reviews, like art, are whatever you can get away with. Or, similarly, one could argue that they had no business evaluating chocolate because its qualities are so subjective it adds fuel to the fire that suggests reviews are no more than one man's opinion. I would disagree.

Donald E. Igelsrud began teaching biology at Delaware Valley College in 1966, became Biology Laboratory director at Northwestern University in 1973, and taught at the University of Calgary from 1976 to 1984. He is founder of ABLE (The Association for Biology Laboratory Education). Currently developing a series of biology videodiscs, he works through his consulting firm: LIFE Consultants, P.O. Box 3097, Postal Station B, Calgary, Alberta, Canada T2M 4L6. His main interests are in increasing awareness and understanding of living phenomena and in developing cooperation among biology teachers and institutions.

Consumer Reports does state that it rated the milk chocolates on the basis of how strong a chocolate taste was produced. Anyone who has eaten Ghirardelli chocolate understands why it was rated the highest. Many people do not like Ghirardelli chocolate for that reason, and prefer the milk flavor of Cote d'Or at the other end of the spectrum. The problem with the *Consumer Reports* evaluation is not that its ratings are inaccurate but rather that, unless one understands their criteria, they are meaningless. Because strength of chocolate flavor is a simple quality to evaluate, compared to a quality like "it tastes good," the *Consumer Reports* study is a good example of a meaningful review. One may not agree with the ratings, but the information one needs to make an evaluation is readily available so that good choices can be made. What would happen if a wide variety of people were each given a different chocolate bar, asked to review it, and then all the reviews were published? What would happen if some of the chocolate bars were fresher than others? How different is this from the way we now review books, films, proposals, papers and other materials for teaching and research?

Films

Because I have looked for the best films of basic biological phenomena throughout most of my teaching career, and because much of my time the past three years has been spent reviewing films for videodisc projects, I feel I am in a particularly good position to comment on film reviews.

In 1977, I was selected as a film reviewer for AAAS's *Science Books and Films*. The publication uses much the same criteria for choosing film reviewers as it does for book reviewers. Interested persons are asked to indicate their fields of interest and competence. Evidence of appropriate background is required by listing degrees and areas of specialization. A vitae is requested and a statement indicating that deadlines will be met must be agreed to. Other organizations which review films have similar requirements, expecting the person to have demonstrated competence in the field. In general, I do not think publications which produce reviews can be faulted for not trying to find interested and competent professionals.

AAAS reviewers are asked to appraise films by rating them: Highly Recommended, Recommended, Acceptable (no serious errors or deficiencies) and Not Recommended. If the film is not recommended, the reviewer is asked to explain why and a second opinion is requested. Level of difficulty is indicated by marking all appropriate catagories: Nursery or Kindergarten, 1st and 2nd grade, 3rd and 4th grade, 5th and 6th grade, 7th-9th grade, 10th-12th grade, College, General Audience and Advanced college or professional, indicating whether the film would be useful for reference at other levels. Excellent, Good, Fair, Poor or Not Applicable ratings are used to evaluate Accuracy of information, Clarity of purpose, Scope (completeness), Optical techniques used in film, Film quality (photography, color, animation, etc.) and value of the film when compared with other similar titles. Reviewers are instructed to "write an evaluation (approximately 200 words) mentioning the film's merits or demerits, describing the contents of the film as they apply to some particular category of viewer, and indicating its value or usefulness for reference, classroom and/or general awareness." Some organizations which solicit reviews are especially concerned about manuscript details, style and readability.

The main purpose of reviews appears to be to have an expert tell what the product is about, to indicate where it might be useful and to comment on its good and bad aspects, as perceived by the reviewer.

The difficulty with this approach is that it is mainly concerned with giving the review authority, and little effort is made to help the consumer make choices. When one attempts to compare reviews to choose films for preview, the choices become very difficult, because there is usually little to compare. If one has viewed a large number of films and then reads reviews, the problems caused by the reviewers' variability in background and point of view are obvious.

Can we do anything about these problems? I think so. The ideal solution would be to have reviewers who have viewed many films in their area of interest and who have, with the help of a set of standards, developed an understanding of what represents the best and worst levels of accomplishment in relation to these standards. If every reviewer has his own standards, it is impossible to make comparisons.

When I taught at Northwestern University in 1974, I showed the film "The Riddle of Photosynthesis" to three colleagues to determine if it would be shown in the introductory course. The following is the description of the film from the new edition (1986) of *educational film/video locator of the consortium of university film centers and r.r. bowker.*

> "Visits the Atomic Energy Commission's Lawrence Radiation Laboratory at Berkeley, California, to explain how the discovery of radiocarbon has contributed to scientific research on photosynthesis. Rebus style formulas, charts, animation, and photographed experiments illustrate the process by which plants transform the sun's energy into usable chemical energy. Algae are submitted to a series of different light exposures to determine chemical change in relation to time of exposure to a radioactive atmosphere, and the new compounds created are identified by paper chromatography and radioautography. Steps in each of these processes are demonstrated. (Magic of The Atom Series)."

One biologist, an ecologist who was in his second year of teaching, was outraged with the film because of its propaganda about atomic energy and its pious treatment of Melvin Calvin. For him, it represented one of the worst films. A neurophysiologist had mixed feelings about the film, but generally did not like it. A biochemist who had done work in the Berkeley lab and knew many of the people in the film felt it did about as good a job as possible of explaining Calvin's work and considered it an excellent film.

If each of these people had written reviews of the film, they would come to different conclusions and would have discussed different aspects of the film. Unfortunately, this difference in perspective is seldom considered when reviews are being considered for publication.

Textbooks

An excellent demonstration of this can be seen in the issue of *Science Books and Films* (20:5 May/June 1985) that was devoted to biology textbooks. Thirty-five pre-college textbooks were reviewed by 105 reviewers. Each book was reviewed by a pre-college biology teacher, a science educator and a biologist. Reviewers were chosen from SB&F's pool of reviewers and from those volunteering via a request in *Science*. Many of the reviewers are well known to the biology teaching community, and no one could doubt the reviewers' competence or interest in identifying quality textbooks.

The criteria for evaluation were grouped into two tables, and ratings were measured as poor, fair, ade-

quate, good and excellent. Grade level and general content categories were: difficulty (grade level), objectivity, accuracy, currency, structure and methods of science, inquiry levels (1-confirmation; 2-structured; 3-guided; 4-open) and motivation, meaning relating to the world of the student. Content evaluation was done by rating 10 key conceptual areas listed in this order: systematics, cell theory, energy transformations, heredity, systems, evolution, ecology, behavior, growth/development and germ theory.

A written review commented on major strengths and weaknesses for the intended students. Reviewers were given organizational and pedagogical criteria for this section. Organizational criteria were concerned with "overall organization and coherence of each text; how adequately the concepts addressed represent the scope of the discipline; how well key topics are explicitly related to one another; how well generalizations are supported by evidence; how well controversial issues are approached with scientific objectivity; and how well questions focus on key concepts. Pedagogical criteria were concerned with "the comprehensibility of the text for the intended student audience; the correctness of the use and definition of technical words and the appropriateness and necessity of their use; how well student participation is encouraged through inquiry, investigation, and experimentation; the adequacy of chapter summaries; how well study questions test for understanding of key concepts; and the extent to which illustrations integrate and/or extend text ideas."

If the rating system and some of the general criteria sound like those listed above for evaluating films, this is no coincidence. These are general criteria used by SB&F and were supplemented with additional criteria proposed by 16 project consultants. Fourteen of these consultants wrote reviews. Of these, nine are described as science educators and their interest in inquiry is shown in the criteria that is used in the tables by accounting for 40 percent of the general ratings.

Despite the fact the reviews' purpose was to "assist individuals and groups involved in the selection of pre-college biology texts" the results are described as follows: "As is evident from the evaluations, what is judged adequate scientific content by one reviewer is not by another. As to pedagogical soundness, there are several theories about that out there, many reflected in these reviews. This is appropriate, for one of the project's main goals is to provide those engaged in text selection with a variety of professional judgements and opinions as well as specific ratings of texts within the structure of objective criteria."

Kathleen Johnson, editor of SB&F, points out that "unbelievably, no one —no single institution in the sciences or education—has attempted a comprehensive evaluation of pre-college biology texts in recent years" and that in their review, "you'll find a richness of commentary that, at the least, makes for interesting reading and that, at the most, will aid you and your school system in the vitally important task of selecting biology teaching materials."

The project is a major contribution to biology education and deserves much praise. I certainly was eager to read the reviews and enjoyed hearing what my colleagues had to say about the books with which I am most familiar. I strongly urge anyone who has not seen it to find a copy; I know they will enjoy reading it.

After reading about this project and realizing the tremendous amount of work that must have gone into producing it, one has to ask how much will it help teachers select and rate textbooks. How easy is it to make comparisons? How valid are the reviews? Can the report do more than make people aware of books they may not have considered? There is no doubt the reviews raise important questions.

When one straightforward item is described by three reviewers in very different ways, one has to question whether the reviewers are using the same standards. For example, the quality of illustrations in John Kimball's *Biology* are described by a teacher as, "Pictures are poorly placed, many are of poor quality, a few are out of date, and color is used as if it were an afterthought;" by an educator as, "The illustrations are greatly improved over previous editions and are valuable adjuncts to the written text;" and by a biologist as, "One of the text's outstanding features is excellent illustration, including line drawings, color photographs (for the first time), and extensive black-and-white photographs. The diagrams present concepts in a very clear and concise manner, helping make this a very readable text."

Differences in point of view are expected and should be considered in developing criteria. No one will be surprised that the most popular biology textbook over the last half century, *Modern Biology*, is described by a teacher in this way, "Anyone seeking a comprehensive biology textbook for use in an introductory high school course need look no further than this edition of *Modern Biology*," or that John Ransom says he would not use it. Ransom states that *Modern Biology* is "probably one of the best of the very *traditional biology* texts available," but rates it poor to good in the general content evaluation table, while it is rated good to excellent by the teacher. Obviously, Ransom is in a difficult position because of the criteria. He feels it does an excellent job of what it sets out to do, but he doesn't like what it is doing and has to summarize his feelings in one statement, so he rates it adequate.

The criteria in the general content table make comparisons difficult because they do not emphasize the most important qualities—those listed under organization and pedagogy. Because the specific content table contains no information about which organisms are discussed, comparisons are not very meaningful. The major difficulty, however, is that because no standards are given for either reviewers or readers to refer to, comparisons are of only limited value.

Texas

Two of the reviewers, Wayne Moyer and William Mayer, wrote *A Consumer's Guide to Biology Textbooks* in 1985 as well. They were concerned about the effect of creationism on the quality of textbooks adopted in Texas, the largest purchaser of textbooks, and reviewed the 18 books considered for adoption there.

When I first looked at their guide, published by People For the American Way, I thought it was more useful than the AAAS project. As I looked more carefully at both projects while writing this review of reviews, I concluded that both are examples of good scholarship, but that neither is particularly helpful in rating books.

The consumer guide does make general comparisons of content easier by listing percentages of text devoted to each area presented. However, because the authors are primarily concerned with reviewing the quality of coverage about evolution and the validity of scientific knowledge, information about other areas is minimal. They also have a few hobby horses that appear repeatedly, e.g. that texts

describe the specific name as the species name instead of emphasizing that the species name is both the genus and specific name. I'm certain that some people will criticize them for being picky but, like the AAAS report, I highly recommend it and know people will enjoy reading it.

Moyer and Mayer do not have a list of criteria with each review, making comparisons difficult. Their point of view, however, is soon clear to the reader and some of their responses are predictable. They also point out important subtleties not seen in some of the more important AAAS reviews. For example, their review of *Modern Biology* points out many small changes that were made to respond to the creationists that were not identified by the AAAS reviewers.

Because neither project rated the textbooks, the validity of the reviews is difficult to ascertain, however one comparison is possible. The ratings in the SB&F's tables seem to have fairly close correlation. By counting the number of excellent ratings in the specific content table, a hierarchy can be produced. The seven highest rated textbooks are listed below with the number of points each received out of a possible 30:

Oram: *Biology: Living Systems*, 1983/ Merrill 16-1/2*
Kormondy: *Addison-Wesley Biology*, 1984 16
Gottfried: *Prentice-Hall Biology*, 1983 13
Otto: *Modern Biology*, 1985/Holt, Rinehart and Winston 13
Haynes: *BSCS Green Version* 1982/ Houghton Mifflin 11
Wallace: *Biology: The Science of Life*, 1981/Scott, Foresman 11
Creager: *Macmillan Biology*, 1985 10
* This rating is based on the 11-point rating, which did not include a teacher evaluation because it did not make the publication's deadline.

Moyer and Mayer's study included ratings by Bassett Maguire, a University of Texas biologist who wrote one of the SB&F reviews, and Steven Schaferman, a geologist and president of the Texas Council for Science Education. They ranked the 18 textbooks, but divided them into general biology, academic biology and advanced biology categories. All of the above books fit into the academic biology category except Wallace, and it was not considered for adoption in Texas.

The three advanced biology textbooks that were considered are: *BSCS Experiments and Ideas* (16/5), *Biology* by Arms and Camp (9/9) and *Biology* by Kimball (10/7). All three were given the highest ratings by Maguire and Schaferman. Their rating in SB&F's is shown in parenthesis (ranks/points).

In the general biology category, six books were evaluated: *Biology and Human Progress* by Tanzer (21/3), *Scott, Foresman Life Science* by Balzer & others (21/3), *Living Things* by Teter & others (17/4), *Biology: An Everyday Experience* by Kaskel & others (21/3), *Biology for Living* by Watkins & others (Silver Burdett did not submit books for review) and *Biology: The Key Ideas* by Wong and Dolmatz (30/0).

None of the books received much support, and all agreed that only Tanzer was acceptable. Moyer and Mayer felt *Biology: The Key Ideas* was "totally unacceptable," which is consistent with its SB&F's rating but possibly overstated. The reviews describe the book as being inadequate for average students; the teacher indicated it would be acceptable for slow learners. I wonder if books which try to simplify biology can ever do well in general review procedures unless special criteria are developed for their consideration.

Nine books were considered in the academic biology category: *BSCS: Green Version* (5/11), *Macmillan Biology* (7/10), *Addison-Wesley Biology* (1/16), *Scott, Foresman Biology* (23/2), *Heath Biology* (10/7), *Biology: Living Systems* (1/16-1/2), *Modern Biology* (2/13), *Prentice-Hall Biology* (2/13) and *Experiences in Biology* by Bauer and others (26/1). Bassett Maguire ranked them as follows: top four in order of quality are BSCS, Macmillan, Addison-Wesley and Scott, Foresman. Of "distinctly lower quality" are Living Systems, Modern Biology and Heath. "Decidedly inferior" are Experiences and Prentice-Hall.

Schaferman's evaluation was similar but clearly independent. Recommended with no reservations are BSCS and Macmillan. Recommended as "marginal" are Scott, Foresman, Addison-Wesley and Heath. "Unacceptable" are Modern Biology and Living Systems. "Totally unacceptable" are Prentice-Hall and Experiences.

This comparison demonstrates that there is a need for standards. Next time I'll talk about how reviews can be improved.