Technical standards and performance standards in the pursuit of quality

RONALD E. ANDERSON

Department of Sociology, University of Minnesota, Minneapolis, Minnesota 55455

The concept of standards lacks clarity because it is applied both to concrete specifications and to the assessment of quality. Confusion is reduced by introducing the notion of interchange expectancy, distinguishing transfer, coupling, and market success as components. Transfer expectancy depends upon the products to be transported or exchanged, but when there is a need to estimate transfer expectancy, a combination of technical standards and performance standards must be applied. Though performance standards are intended to produce quality, the quality of the instrumentation to measure performance standards is often neglected. The development of high-quality performance standards first requires the delineation of the relevant dimensions of performance quality. The process of giving precision to performance standards may lead to the emergence of more acceptable technical standards. To illustrate some of these points, data is presented from the CONDUIT system for quality evaluation of computer-based curriculum materials. The performance quality of these products includes the dimension of transfer expectancy, as well as educational and scientific merit. Increased specification and impact of technical standards may provide greater incentives to authors and hence lead to the invention of products which have higher educational and scientific quality.

The relationship between technical standards and performance standards is both confusing and controversial. It is confusing in that often the two types of standards are mistakenly intertwined, and it is controversial because some take the vehement position that a negative association exists between performance quality and level of standardization. The strongest argument against technical standards is that standards suppress the variation upon which quality often depends and they inhibit innovation and creative change. Another major argument is that technological environments, especially computer systems, are dynamic and changing rapidly, and, therefore, standards are considered to be unfeasible, irrelevant, or too costly. The proponents of standards, on the other hand, promote standards because they are practical, cost effective, and foster sharing and intercommunication. Unfortunately, the debate tends to separate the inventors from the producers, the creators from the consumers, and the artists from the scientists. The opponents of standards claim a value orientation of creativity, while the supporters argue for the importance of application and production. This dualism is so pervasive that the word "standardization" has become a negative symbol in contemporary society.

Rather than continue this unresolvable debate, it

The preparation of this report was supported in part by CONDUIT, a National Science Foundation funded project to study transfer of computing ideas. Special thanks are due James Johnson, Trinka Dunnagan, John Castellan, and David Kniefel for their assistance.

would be more useful to try to clarify the issues in greater detail and, in particular, to explore the relationship between standards and quality. To begin this exercise, the following terms should be defined:

Technical standards are specifications which provide a model or template for determining whether or not a particular situation is constructed within a region of acceptability. Technical standards are a direct function of cooperative consensus and their impact can often be measured by the number of adoptions. The criteria for evaluating whether or not a particular situation conforms to the standard are straightforward and precise; the outcome of the evaluation is either a yes/no or true/false statement or a precise estimate of error. Even if the standard is highly complex and extensive, as in a programming language standard, the adequacy of the situation can be determined by counting or identifying each deviation from the standard.

Performance standards are rules or principles for the assessment of relative merit or quality. Performance standards do not always depend upon broad consensus and acceptance; in fact, such acceptance is very difficult to achieve because of the ambiguity of most performance standards. Applying a performance standard to a particular situation typically results in an imprecise statement of relative quality. Performance standards arise from, and closely parallel, objectives and values.

Another important dimension of standardization is the degree to which standards are legitimated by outside authority. *De facto* standards are the dominant practices and common agreements of a

Figure 1. Differences between technical and performance standards in relation to criteria.

community. These norms depend solely upon social consensus and are sometimes highly pervasive with complete conformity to a rule. In contrast, *formal* standards receive the backing of an established authority for the maintenance of standards. The formalization of technical standards tends to occur at the interorganizational level, e.g., American National Standards Institute, whereas formalization of performance standards tends to emerge within particular organizations, e.g., a university may establish a minimum GPA for entrance.

Even though it is important to distinguish between technical standards and performance standards, there are basic similarities. The fact that both are fundamentally concerned with quality is perhaps the key to why we tend to confuse the two. Performance standards are directly concerned with quality, while technical standards arise from the goals of improving the quality of interchanges. The technical standard itself may only implicitly involve quality judgments, but it is often much more explicit in the case of a complex standard such as a programming language standard. For such large, complex areas of standardization as programming languages, it is inevitable that even a good standard such as ANSI FORTRAN, will not encompass the total range of the language. For instance, the ANSI FORTRAN does not deal directly with file interfaces and character

string manipulation. Also, the proposed BASIC standard omits perhaps as much as 50% of the language facilities required by a complex BASIC program. It becomes possible, in fact necessary, to estimate the degree to which a particular program conforms to the standard. More abstractly, it becomes necessary to evaluate the situation in terms of "reasonableness" or "transferability," and, in some cases, it may be useful to establish a performance standard with respect to transferability. Although it sometimes is possible to construct checklists and fairly exact estimates of the percentages of conformity with a language standard, it may be necessary to resort to relatively less precise techniques such as peer ratings. The foregoing analysis demonstrates the dilemma that complex technical standards, even though they may be explicit, inevitably are muddled by judgments of quality and the need to superimpose performance standards on technical standards.

Both technical standards and performance standards are closely associated with criteria which are the basis or the procedures for relating standards to concrete situations. There are several aspects of a criterion, as illustrated in Figure 1. First, there is the continuum along which variation in items can occur, e.g., size, in the case of technical standards, or efficiency, in the case of performance standards. Secondly, there is a point on this scale that is specified

by the standard as desired or optimal. Thirdly, there is a range of acceptability on the continuum. The standard either explicitly or implicitly specifies this range, but, in the case of technical standards, the range is likely to be interpreted as tolerable error variation, whereas with performance standards the range is the zone between the maximum and minimum quality, or between the most and least acceptable states. These distinctions are important because otherwise almost no effective difference exists between technical standards and performance standards.

Some of the semantic confusion over standards results from implicitly conceptualizing "standard" as equivalent to only one aspect of a criterion. Sometimes the term "standard" is used to refer to the underlying dimension or continuum; sometimes it refers to the state corresponding to a specific point on the continuum; and sometimes it refers to the range of acceptable variation along the continuum. Interestingly, these different conceptualizations of standards are nearly equivalent to the alternative conceptualizations of "value" within the social science. As pointed out by Williams (1968), there are two major orientations toward the concept of value: values as criteria and values as states or end states. In other words, these two somewhat conflicting points of view are based upon the definition of values as underlying, continuum vs. values as states or points on a continuum. The third aspect of criteria, the range of acceptable variation, is more commonly found as a conceptual approach in more practical contexts such as the assessment of medical care (Brooks, 1973). The close association between standards and values is not accidental because standards arise from and are dictated by values and objectives. The important point is that all are primarily concerned with quality. The measurement of quality using performance standards is a very general problem encompassing such diverse areas as evaluation research, technological assessment, quality evaluation and control, accountability review, performance measurement, value measurement, etc. Some quality assessments concentrate upon process, while others emphasize outcome; the evaluation of immediate states vs. future impacts are alternative assessment foci. Research on procedures for assessment of the quality of medical care has demonstrated that it is essential to include both process factors and outcome factors (Brooks, 1973). As the evaluation of quality in any given area matures and develops, the dimensions of quality become more clearly delineated (cf. McTavish & Anderson, 1973; Noble, 1974). In the area of computer technology, it is increasingly apparent that interchange or transport is a major outcome dimension of quality. As a computing community, such as the laboratory automation or instructional computing community, matures and

refines its techniques, it apparently becomes increasingly convinced of the importance of interchange or transport as a key dimension of quality (see Castellan, 1973). Interchange is an outcome variable, yet it is a factor that requires assessment prior to actual transport. Thus the transfer expectancy of an item often deserves to be evaluated, even though its estimation may be, at best, speculative.

TRANSFER EXPECTANCY

Interchange expectancy necessarily depends upon the objects to be moved and the process required to move them. Some products, e.g., Teletype terminals, primarily pose problems of interconnectability or coupling, while other products, e.g., packages such as SPSS, primarily pose problems of transport and installation. Another type of product (e.g., a book on computer science for social scientists) is very simple to transfer but ultimately finds few users or consumers. Thus, for these three types of products, there are three corresponding domains of concern: coupling expectancy, transfer expectancy, and market expectancy. Those persons or agencies dealing with hardware are generally preoccupied with coupling expectancy, those in the software business tend to be concerned with transfer expectancy, and those persons or agencies in the delivery system business, e.g., libraries, computer centers, clearing houses, are largely concerned with market expectancy.

Problems of coupling lend themselves to solution by the establishment of technical standards. This is less true for transfer problems and even less so for market problems. Thus, performance standards are often associated with market expectancy and technical standards with coupling expectancy. Both types of standards are critical to issues of transfer expectancy. For instance, the transport of computer-based curriculum materials depends heavily upon both technical and performance standards. A delivery system for these products must also investigate coupling and market expectancy as well as transfer expectancy. Its success depends in large part upon the marketplace which is constrained by idiosyncratic local environments with unique coupling and implementation problems. From the standpoint of a delivery system, investment in the transport of a product must depend upon the expected outcomes of transfer. Such assessments can only be obtained by expert judgments of quality, i.e., transfer quality.

Although evaluation of transfer expectancy is critical to a system of technological interchange, the quality of the available instrumentation for such evaluation is undeveloped and usually neglected. Consistent application of performance standards in this type of area can have the side benefit of fostering the development of appropriate technical standards.

Table 1
Disciplinary Source of Packages Reviewed

Social Science	
Social Sciences	26%
Business	21%
Economics	5%
Physical Science	
Chemistry	30%
Math	9%
Physics	9%
	<hr/> 100%

Table 2
Types of Packages Reviewed

Simulations	17%
Games	9%
Canned Analytic Tool	23%
Courseware Module	29%
Integrated Text Module Package	11%
Data Base Package	11%
	<hr/> 100%

For instance, CONDUIT's (Dunnagan, 1976) attempt to systematically evaluate the transfer expectancy of computer programs has contributed to the ongoing project to produce a formal BASIC language standard. The CONDUIT project developed a procedure to systematize the estimation of transfer and market expectancies. The system relies upon the peer review process with particular questions directed to expert referees. Analysis of these review ratings of curriculum packages is presented here to illustrate some measurement problems and potential solutions. The data also include findings pertinent to the complex interrelationship between scientific merit and standardization.

A SYSTEM FOR QUALITY ASSESSMENT OF COMPUTER-BASED CURRICULUM MATERIALS

Recognizing the need to obtain the advice of scientists, teachers, and computer specialists in evaluating the quality of educational materials, CONDUIT established a peer review system like that of professional journals and publishing houses. During the 1973-1974 period, over 150 reviews were obtained from 97 discipline experts and instructors, some with prior experience using the package under review in their courses. Eighty packages were reviewed and distributed across six disciplines (see Table 1). Despite the wide diversity in the nature of these packages (see Table 2), most included a computer program, operating instructions for the student, and pedagogical suggestions for the instructor. In a few instances, the packages consisted primarily of a machine-readable data base (see Table 2). Each

reviewer was asked to write a brief summary critique of the package and to rate the package on 6-point scales of good/poor quality. (Rating scale items are listed in Tables 5 and 6.) After rating the package on each of 21 separate items, reviewers gave a general rating of the overall package.

Several questions can be explored with the data: (1) Is such a systematic evaluation of quality possible across a wide variety of raters, disciplines, courses, and packages? Relatively high levels of consensus among raters would tend to validate the rating technique. (2) What are the dominant quality orientations used in evaluating these computer packages? What are the major sets or clusters of criteria seen as relevant to these packages? (3) Are there major differences in perspectives and rating tendencies, depending upon the background of reviewers? Course experience with the package and disciplinary identification can be contrasted to investigate this question. (4) What is the interrelationship between "transportability" as a quality criteria, and other dimensions of quality? (Correlations between transportability ratings and other considerations pertain to this question.)

The first question of evaluative consensus or agreement can be investigated by looking at the variance in ratings comparing the within variance with the between variance. This analysis is presented in Table 3 for the overall package rating. The between variance is significantly greater ($p < .001$) than the within-package variance. An elaboration of this contrast is presented in Table 4, where the ratings have been dichotomized into either positive or negative evaluations. (A 1-3 rating is negative and a 4-6 rating is positive.) For all 80 packages, there are only nine disagreements (11%) across this dichotomy.

Table 3
**Analysis of Variance of Overall Package
Rating for 150 Package Reviews**

	MS (Mean Square)	df
Between Packages	1.57*	79
Within Packages	0.75	71

* $F = 2.11, p < .001$

Table 4
Consensus on Overall Package Ratings*

	Number of Packages	Percentage
Agreement on		
Positive Evaluation	63	79
Agreement on		
Negative Evaluation	8	10
Disagreement	9	11
	<hr/> 80	<hr/> 100

*Consensus is assumed if a package receives all good/adequate ratings or all poor/inadequate ratings. If the ratings are mixed, then dissensus or agreement is assumed.

Table 5
Evaluation Criteria Contrasting Physical and Social Science Raters

	Correlation of Each Criterion with Overall Evaluation		Mean Ratings of Importance	
	Physical Science	Social Science	Physical Science	Social Science
1. Definition of main concepts	.46	.57	3.8	4.2
2. Discussion of basic premises and assumptions	.36	.61	3.7	4.1
3. Specification of internal workings	.48	.29	2.8	2.1**
4. Sufficiency and merit of substantive content	.12	.64	3.7	3.7
5. Discussion of literature germane to the substantive issues	.60	.53	1.7	3.4**
6. Readability, clarity of substantive documentation	.38	.47	4.0	3.9
7. Effectiveness of student instruction in theories and principles	.49	.62	3.3	3.8
8. Effectiveness of student instruction on how to apply theories	.63	.47	3.4	3.8
9. Potential for improving instructor ability to communicate theories	.38	.40	3.5	3.7
10. Potential for improving instructor ability to communicate method	.68	.23	2.8	3.5*
11. Educational value of this approach to teaching concepts	.72	.58	3.7	3.8
12. Educational value of this approach to teaching method	.54	.41	3.2	3.3
13. Variability and flexibility in response to possible student use	.59	.25	3.1	3.3
14. Challenge to student creativity	.40	.31	3.1	3.8**
15. Transportability of the computer program	.38	.08	4.0	3.8
16. Manageability in the classroom	.32	.50	3.8	4.2
17. Utility for self-initiated student work (independent study)	.65	.10	3.1	2.8
18. Adequacy of directions for student use without explication	.23	.26	3.2	3.4
19. Useability by students without previous computer experience	.20	.31	2.9	3.8***
20. Sufficiency of examples showing package use	.44	.54	3.8	3.4
21. Justification for computer rather than noncomputer approach	.49	.56	3.5	2.9**
N (Number of raters)	36	63	36	63
R ² (prediction of overall package rating by 21 evaluation criteria)	.98	.84		

**p* (two-tailed) < .10

***p* (two-tailed) < .05

****p* (two-tailed) < .01

Thus, there is agreement on the direction of the evaluation in 89% of the cases. While some raters complained about the fixed nature of the rating scales and the obtuse wording of some of the criteria, their comments tend to support the adequacy of the instrument as a sensitizing device and as an assessment tool for quality evaluation.

The question of evaluative differences among the raters is investigated by examining the ratings of the importance of each criterion and by intercorrelating the actual package rating on each criterion with the overall package rating. The results are presented in Table 5, contrasting physical science with social science raters, and in Table 6, contrasting raters with course experience with those lacking course experience with the particular package under review. It is noteworthy that social scientists rated "discussion of literature," "challenge to student creativity," and "usability by students without previous computer experience" as more important than did the physical scientists. And instructors with classroom experience rated "manageability in the classroom" as more important but "specification of internal workings" as less important than did those lacking course experience. While there are a few interesting differences such as these, the major finding is that there is substantial uniformity in the perception of the relative importance of evaluation criteria.

The dimensionality of the 21 item ratings was explored with factor analyses on the intercorrelation matrix (Table 7). Due to low measures of sampling adequacy (Kaiser & Rice, 1974), Items 10, 12, 13, and 17 were dropped. A principle components factor analysis with varimax rotation on the remaining items yielded a five-factor solution as follows:

I. Quality of substantive pedagogy
(Items 4,5,7,9,11)

II. Quality of presentation of concepts and theories
(Items 1,2,6,18)

III. Quality of presentation of methods and techniques
(Items 3,8,20,21)

IV. Quality of student orientation
(Items 14,19)

V. Quality of transfer expectancy
(Items 15,16)

These five factors account for 70% of the total variance. Factor scores from these five factors were analyzed using multiple regression in the prediction of the overall package rating; the results are presented in Table 8. The substantive pedagogy dimension of

Table 6
Impact of Course Experience Upon Evaluation Criteria

	Correlation of Each Criterion with Overall Evaluation		Mean Ratings of Importance	
	Course Experience	No Course Experience	Course Experience	No Course Experience
1. Definition of main concepts	.59	.46	4.1	4.0
2. Discussion of basic premises and assumptions	.59	.54	4.0	4.0
3. Specification of internal workings	.35	.09	1.9	2.9***
4. Sufficiency and merit of substantive content	.63	.61	3.7	3.7
5. Discussion of literature germane to the substantive issues	.49	.55	1.8	2.6**
6. Readability, clarity of substantive documentation	.37	.53	4.1	3.8
7. Effectiveness of student instruction in theories and principles	.55	.69	3.7	3.4
8. Effectiveness of student instruction in theories and principles	.36	.49	3.7	3.6
9. Potential for improving instructor ability to communicate theories	.29	.68	3.7	3.6
10. Potential for improving instructor ability to communicate method	.20	.61	3.5	3.0
11. Educational value of this approach to teaching concept	.56	.63	3.6	3.9**
12. Educational value of this approach to teaching method	.43	.51	3.4	3.2
13. Variability and flexibility in response to possible student use	.00	.45	3.2	3.3
14. Challenge to student creativity	.14	.41	3.5	3.5
15. Transportability of the computer program	.16	.38	3.7	4.0
16. Manageability in the classroom	.60	.49	4.2	3.7**
17. Utility for self-initiated student work (independent study)	.06	.31	2.6	3.3
18. Adequacy of directions for student use without explication	.25	.31	3.2	3.5
19. Useability by students without previous computer experience	.17	.42	3.7	3.1
20. Sufficiency of examples showing package use	.63	.30	3.6	3.4
21. Justification for computer rather than noncomputer approach	.53	.56	3.1	3.1
N =	55	39	55	39
R ² =	.83	.86		

p* (two-tailed) < .10*p* (two-tailed) < .05****p* (two-tailed) < .01

Table 7
Intercorrelations Among 21 Criteria of Package Quality (N of raters = 99)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1.	1.00	.69	.29	.35	.44	.56	.45	.28	.34	.15	.31	.23	.01	.13	.18	.35	-.05	.46	.22	.35	.25
2.	.69	1.00	.30	.48	.49	.54	.51	.39	.38	.15	.37	.24	.17	.19	.14	.40	-.03	.38	.29	.25	.23
3.	.29	.30	1.00	.19	.38	.21	.30	.19	.12	.11	.25	.15	.03	.01	.21	.03	-.01	.17	.01	.25	.25
4.	.35	.48	.19	1.00	.46	.54	.61	.45	.52	.33	.57	.30	.20	.22	.06	.31	.03	.18	.23	.17	.25
5.	.44	.49	.38	.46	1.00	.41	.50	.30	.33	.21	.38	.21	-.10	.07	.24	.25	-.07	.33	.09	.27	.21
6.	.56	.54	.21	.54	.41	1.00	.51	.22	.28	.21	.26	.09	.22	.15	.15	.36	.11	.43	.46	.25	.08
7.	.45	.51	.30	.61	.50	.51	1.00	.54	.56	.33	.66	.27	.19	.16	.21	.49	.05	.22	.26	.26	.26
8.	.28	.39	.19	.45	.30	.22	.54	1.00	.27	.29	.33	.39	.10	.21	.08	.21	-.01	.23	.13	.33	.28
9.	.34	.38	.12	.52	.33	.28	.56	.27	1.00	.43	.66	.28	.24	.31	.17	.31	.00	.19	.26	.14	.35
10.	.15	.15	.11	.33	.21	.21	.33	.29	.43	1.00	.33	.81	.13	.15	.26	.17	.06	.17	.26	.24	.30
11.	.31	.37	.25	.57	.38	.26	.66	.33	.66	.33	1.00	.38	.08	.35	.25	.30	-.01	.08	.18	.16	.44
12.	.23	.24	.15	.30	.21	.09	.27	.39	.28	.81	.38	1.00	.06	.21	.20	.20	-.07	.11	.15	.26	.34
13.	.01	.17	.03	.20	-.10	.22	.19	.10	.24	.13	.08	.06	1.00	.51	-.07	.23	.51	-.02	.21	-.06	.31
14.	.13	.19	.01	.22	.07	.15	.16	.21	.31	.15	.35	.21	.51	1.00	.09	.15	.26	-.08	.15	-.01	.27
15.	.18	.14	.21	.06	.24	.15	.21	.08	.17	.26	.25	.20	-.07	.09	1.00	.44	.04	.19	.14	.32	.16
16.	.35	.40	.03	.31	.25	.36	.49	.21	.31	.17	.30	.20	.23	.15	.44	1.00	.16	.16	.25	.30	.21
17.	-.05	-.03	-.01	.03	-.07	.11	.05	-.01	.00	.06	-.01	-.07	.51	.26	.04	.16	1.00	-.02	.05	-.03	.16
18.	.46	.38	.17	.18	.33	.43	.22	.23	.19	.17	.08	.11	-.02	-.08	.19	.16	-.02	1.00	.36	.30	.14
19.	.22	.29	.01	.23	.09	.46	.26	.13	.26	.26	.18	.15	.21	.15	.14	.25	.05	.36	1.00	.24	.26
20.	.35	.25	.25	.17	.27	.25	.26	.33	.14	.24	.16	.26	-.06	-.01	.32	.30	-.03	.30	.24	1.00	.35
21.	.25	.23	.25	.25	.21	.08	.26	.28	.35	.30	.44	.34	.31	.27	.16	.21	.16	.14	.26	.35	1.00

package quality clearly dominates the assessment of overall quality while student orientation and transfer expectancy are not significant components of the overall evaluation. It would appear that the raters have separated the dimension of transfer expectancy from what are deemed relevant quality dimensions.

To adequately interpret these results, it is necessary to consider the reviewers' role set. In the design and implementation of computer-based curriculum materials, there are five major roles, each with its own dominant quality orientation: (1) Scientist, scholar—quality of ideas and substantive content. (2) Author—quality of communication through written materials. (3) Teacher—quality of pedagogy. (4) Student—quality of mental stimulation and learning. (5) Computer specialist—quality of interchange and computing integrity. The reviewer typically occupies the teacher, scientist, and author roles and avoids the technical computer details. The perspective from this role system probably leads to a belief that transfer expectancy need not be of major concern. Selection of reviewers who are typical academics will result in a bias toward the pedagogical, scholastic, and literary dimensions of quality.

In addition to the role orientation of the evaluator, the structure and content of the rating instrument affects the evaluations. The emphasis in the list of criteria upon substantive and pedagogical issues undoubtedly structures and perhaps biases the evaluation task. Expanding the issue of transfer expectancy from 1 or 2 items to 10 or 15 might shift the system of quality assessment dramatically. Taking advantage of the foregoing analysis, CONDUIT has produced a revised rating instrument. The new set of evaluation criteria directly take into account the five roles described above by including several items central to each role. Future work by CONDUIT may result in another rating instrument which concentrates totally upon technical criteria including such items as: conformity to ANSI language standards, description of linkage to operating system, inclusion of checks for error conditions, structural modularity of code, accuracy of numeric methods, etc. Such systematic assessments of technical quality would contribute importantly toward our understanding of transfer expectancy.

CONCLUSION

The original issue of whether or not quality and standardization are truly inversely related should be rephrased to: Is there a low or negative association between scientific quality and transfer quality? While we cannot provide definitive answers to this question,

Table 8
Multiple Regression of Factor Scores on Overall
Package Rating (N of Raters = 99)

Factor Scores	Standardized Beta Values	Simple Correlation
F1 - Substantive Pedagogy	.40*	.73
F2 - Presentation of Concepts and Theories	.16*	.61
F3 - Presentation of Methods and Techniques	.26*	.63
F4 - Student Orientation	.12	.43
F5 - Transfer Expectancy	.10	.48

* $p < .01$ (multiple $R = .82$, $R^2 = .67$)

our data show that academic reviewers are likely to perceive transfer expectancy to be independent of the overall quality of computer materials. Perhaps this is due to the alienation from computer techniques which most academics experience. In any case, the belief does result in part from the particular role perspective of the evaluator.

Taking a broader perspective, we prefer to retain our belief in the proposition that high scientific and intellectual quality are in part a function of high transfer expectancy. Increasing the pervasiveness of technical standards and the potential for interchange may actually provide greater incentives to authors and hence lead to the invention of products which in fact have higher educational and scientific merit. Lack of incentives due to the perceived novelty and rapidly changing character of the computing field seems to inhibit stable and high-quality development.

REFERENCES

- BROOKS, R. H. *Quality of care assessment: A comparison of five methods of peer review*. DHEW Publication No. HRA-74-3100, 1973.
- CASTELLAN, N. J., JR. Laboratory programming languages and standardization. *Behavior Research Methods & Instrumentation*, 1973, 5, 249-252.
- CASTELLAN, N. J., JR. Standards: Real-time extensions to programming systems. *Behavior Research Methods & Instrumentation*, 1976, 8, 207-210.
- DUNNAGAN, T. CONDUIT standards and procedures: A practical experience. *Behavior Research Methods & Instrumentation*, 1976, 8, 205-206.
- KAISER, H., & RICE, J. Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 1974, 34, 111-117.
- MCTAVISH, D., & ANDERSON, R. E. *Evaluation tools for methodological characteristics of social research*. Technical Report, Minnesota Systems Research, Inc., Minneapolis, Minn., 1973.
- NOBLE, J. H., JR. Peer review: Quality control of applied social research. *Science*, Sep 1974, 916-920.
- WILLIAMS, R. M. Values. *Encyclopedia of the social sciences* (2nd ed.), 1968, 283-291.