DOCUMENT RESUME

ED 039 127 SE 008 171

AUTHOR Strauss, Samuel
TITLE An Analysis of Published Evaluations of BSCS.
INSTITUTION Johns Hopkins Univ., Baltimore, Md.
PUB DATE 8 Mar 70
NOTE 17p.; Paper Presented at Annual Meeting of the National Association for Research in Science Teaching (43rd, Minneapolis, Minne., March 5-8, 1970)

EDRS PRICE EDRS Price MF-$0.25 HC-$0.95
DESCRIPTORS *Biology, *Curriculum Evaluation, Educational Research, *Evaluation, *Research Criteria, *Research Methodology, Secondary School Science
IDENTIFIERS Biological Sciences Curriculum Study

ABSTRACT

To illustrate a system for analyzing research reports, four published evaluations of the Biological Sciences Curriculum Study (BSCS) program were analyzed in terms of problem raised, previous work cited, objectives stated, hypotheses formulated, assumptions made, population studied, sample drawn, instruments used, design examined, procedure followed, safeguards taken, observations recorded, findings assembled, statistics interpreted, interpretations discussed, conclusions reached, limitations recognized, further work projected, improvements suggested and clarity of report. The analysis is reported as a chart with each aspect of each report graded from A to E according to the author's judgment of the strength of the study in that area. The author concludes that the case for BSCS has yet to be proved. (EB)

AN ANALYSIS OF PUBLISHED EVALUATIONS OF BSCS

Samuel Strauss
The Johns Hopkins University, Baltimore, Maryland 21218

Presented at the meeting of the

NATIONAL ASSOCIATION FOR RESEARCH IN SCIENCE TEACHING

Minneapolis, Minnesota
March 8, 1970

SAMUEL STRAUSS
9308 ADELPHI ROAD
ADELPHI, MARYLAND 20783

Samuel Strauss
Johns Hopkins
March 1970

AN ANALYSIS OF PUBLISHED EVALUATIONS OF BSCS

Studying reports of research projects can be quite frustrating. What was the investigator really doing? How well did he standardize his instruments, sample his subjects, assemble his data? Where did he introduce safeguards to protect the integrity of his study? Why are his conclusions different from those of other workers? Was the work itself sound, but the reporting inadequate?

Even more confusing is the task of summarizing a number of studies in the same area. Locating the reports is a problem -- though ERIC (1) is now a help -- but more difficult is reducing the published papers to some common bases so that their findings can be compared. The usual compromise is to accept the conclusions of the authors as stated and let it go at that; but what if the conclusions are unwarranted -- 38 per cent in one study (8) -- or if the projects are so different as to make comparisons almost meaningless? These questions face all consumers and reviewers of research and, because they are so complex, leave those who would like to utilize or explain research findings more or less at sea in the middle of research.

To help bring some order into the processes of evaluating and reviewing published studies, I have developed a fairly simple set of Guidelines, described elsewhere (7), for analyzing research reports. (A reprint of this paper is attached.) These Guidelines, which are easy to apply and which focus emphasis on the chief qualities of good research, are comprised of the following twenty criteria:

1. Problem raised, 2. Previous work cited, 3. Objectives stated, 4. Hypotheses formulated, 5. Assumptions made, 6. Population studied, 7. Sample drawn, 8. Instruments used, 9. Design examined, 10. Procedures followed, 11. Safeguards taken, 12. Observations recorded, 13. Findings assembled, 14. Statistics interpreted, 15. Interpretations discussed, 16. Conclusions reached, 17. Limitations recognized, 18. Further work projected, 19. Improvements suggested, 20. Clarity of report.

The above criteria -- which are also useful for analyzing individual reports, for planning projects and for writing proposals -- can be used to compare and review a number of reported studies within a given field. Such an analysis, in the form of a chart, of the known assessments of the three BSCS versions in high school biology, constitutes the body of this paper. A literature search has revealed only four published summative evaluations, although there are some unpublished dissertations (2,3,4) and some published reports which use BSCS materials in dealing with other research topics (examples: 5, 6). Each criterion above is used in the chart to concisely and critically describe some aspect of each study, with enough information to yield a fairly adequate resume of the entire published evaluation.

Each aspect of each report is also graded to indicate how well, in my own subjective opinion, each criterion has been met, according to the following scale: A, for a complete and clear statement fully satisfying the criterion; B, for a fairly good statement, but lacking some essential quality; C, for a weak statement, or a strong implication somewhere in the report; D, for

a quite inadequate or confusing statement, and E, for the lack of either a statement or an implication. It would be wrong to transform the letter grades to numerical values in order to compute a total score or a mean score, since all the twenty criteria do not have equal weight.

To be completely fair to the authors of the papers analyzed, the chart has been submitted to them for their comments and has been partly revised in the light of their criticisms. It is also fair to point out that published papers are sometimes altered or shortened by editors.

Although the chief purpose of this chart is to illustrate how the Guidelines may be applied, it is also possible to draw from it some conclusions about the subject matter, that is, the results of the evaluations of the BSCS curricula. My own general opinion, after this review, is that the case for BSCS has yet to be proved. That is a pity. After hundreds of competent and enthusiastic people have spent a vast number of man-hours in developing what is obviously a fresh, bright, complete and up-to-date series of curricula for secondary school biology, it seems a shame that more conclusive published evidence is not yet available as to its validity -- that is, sound proof that the BSCS program truly accomplishes what it sets out to do. Good evaluation strategy (9) would seem to require proper sampling (10) if the findings are to be honestly generalized, carefully standardized assessments of growth and achievement, control groups and other design safeguards. After all, in evaluating the effects of science teaching, the emphasis should be on scientific rigor.

PUBLISHED REPORT	1. PROBLEM RAISED	
Grobman, H., Wallace, W.L. and Klinkmann, E. "The BSCS 1961-62 Evaluation Program." (in four papers) BSCS Newsletter, No. 19, September 1963.	Implied: How effective was the BSCS program in the three versions, and also when compared to non-BSCS biology, in its trial use during 1961-62?	C
Lisonbee, L. and Fullerton, B.J. "The Comparative Effects of BSCS and Traditional Biology on Student Achievement." School Science and Mathematics, 64:594-598; October 1964.	Would high, middle and low ability students from different schools enrolled in the BSCS program do as well on achievement tests as those in traditional biology classes?	A
George, K.D. "The Effects of BSCS and Conventional Biology on Critical Thinking." Journal of Research in Science Teaching, 3:293-299; 4, 1965.	Would the BSCS program be more successful than conventional biology in developing critical thinking ability?	A
Staff of Psychological Corp. "A Report of the BSCS End-of-Year Evaluation Program, 1964-1965." BSCS Newsletter, No. 30, January 1967.	Implied: How would students perform on the different forms of the different BSCS achievement tests, and how would this be related to academic ability and to reading skills?	C

	2. PREVIOUS WORK CITED	3. OBJECTIVE STATED	
E	Only two encyclopedic books.	Implied: To justify the worth of the BSCS program.	C
A	Several references about the development of this (then) new program.	To measure the effect of BSCS vs traditional biology on student achievement.	A
D	Four old studies in non-science areas, but none on the many papers on scientific thinking, nor on recent reports.	To test the effect of BSCS vs conventional biology on development of critical thinking.	A
E	No references.	Implied: To develop good standardized tests for the BSCS program.	C

4. HYPOTHESES FORMULATED	5. ASSUMPTIONS MADE
Implied in the tabulation of findings: There would be no significant differences on three achievement post-tests and three attitude post-measures between: male, female 10th graders taught with or without lab blocks; with Blue, Green or Yellow versions -- compared with male, female control groups using non-BSCS materials,* and with 9th graders using the three versions. C	None stated. E
Clear statement of eight hypotheses: There would be no significant differences in achievement, with CTMM and ITED scores held constant, on the Nelson Biology Test or the BSCS Comprehensive Final, between BSCS vs traditional classes of high, middle and low ability, nor among students in different schools. A	None stated. E
Clear statement of four hypotheses: There would be no significant differences in critical thinking ability between pupils taught by Blue, Green or Yellow versions vs pupils taught by conventional biology;* nor among pupils taught by the three versions. A	Listed 8 assumptions, and defended 7 of them by citations and evidence. A
Implied in the tabulation of findings: There would be no significant differences in scores of 10th graders on two parallel forms of each of the three versions' Quarterly Achievement Tests, nor on two parallel forms of BSCS Comprehensive Final, when compared by: sex, academic ability, BSCS version or reading ability. C	None stated. E

*No other information on this item

6. POPULATION STUDIED	8. SAMPLE DRAWN
Near 37 centers in the U.S., 361 teachers taught 39,000 students, stratified (but not matched) by: grade, sex, lab block or non-block use, BSCS version. Control groups, matched by teacher and school, of 3944 students taught without BSCS materials* by 136 teachers. Academic ability of students above average. B	None; only the 18 sub-populations to which the findings can be said to apply. D
3500 10th grade biology students who were tested for homogeneity by Bartlett's Test.* C	By random sampling chose expt'l and control groups, stratified into high, middle and low ability and by schools.* B
In four suburban Chicago schools, 19 classes ranging from 17 to 25 were taught by 10 volunteer teachers. Classes, teachers and facilities were found equivalent by a questionnaire.* Academic ability of students above average. B	"No claim is made as to the representativess of the biology classes." D
9,846 10th graders took achievement tests. 998 took Davis Test.* 907 took Illinois Test.* "Among versions, participating schools were similar as to type of community, and as to type, size and facilities of schools." Teachers had similar education, experience and work loads.* "Statistical information may be found in the BSCS Manual" but not in this report. C	Sampling hinted at.* E

* No other information on this item.

8. INSTRUMENTS USED	9. DESIGN EXAMINED	
SCAT (B) for academic ability. BSCS Comprehensive Final and Coop Biology Test (Y) for achievement. Impact Test for reasoning ability and understanding. Three attitude measures from TOUS, Purdue attitude scale and semantic differential. Questionnaire on teacher background and schools.	Independent variable: SCAT scores, groups equated by covariance. Dependent variables: Post-tests on achievement and attitudes (but without pre-tests, no measure of gains).	B C
California Test of Mental Maturity (CTMM) for academic ability. Iowa Test of Educational Development,(6) (ITED) for scholastic ability. Nelson Biology Test and BSCS Comprehensive Final for achievement.	Independent variables: CTMM and ITED scores. Dependent variables: Post-tests on achievement (but without pre-tests, no measure of gains).	B C
Otis Quick-Scoring Test (Gamma Fm) for academic ability. Watson-Glaser Critical Thinking Appraisal (Rev. Zm).	Independent variables: IQ scores, equated by covariance, and pre-test scores on W-G. Dependent variable: Post-test scores on W-G.	B B
DAT (L) for academic ability. BSCS Quarterly Achievement Tests (R and S) for each BSCS version. BSCS Comprehensive Final (I and II). Davis Reading Test (2A and 2D) and the Illinois Natural Science Reading Comprehension Test, for reading skills.	Independent variables: DAT scores and pre-tests on reading. Dependent variables: Two post-tests on reading and six post-tests (but no pre-tests)on achievement.	B C

(Note: None of the above instruments were defended in any report.)

10. PROCEDURES FOLLOWED

All students took SCAT at start of trial year. In 10th grade 29 classes used Blue, 39 Green, 49 Yellow versions with lab blocks; while 57 used Blue, 52 Green, 39 Yellow without blocks. In 9th grade 5 classes used Yellow with blocks; while 11 used
C Blue, 12 Green, 6 Yellow without blocks. 125 classes did not use BSCS materials.* All students took 3 achievement and 3 attitude tests at end of year. BSCS students also took Quarterly Achievement Tests. All teachers filled described questionnaire. (Explained here and there; also implied in tables.)

E No description.
(Full details in dissertation, but unpublished.)

All students took Watson-Glaser as pre-test in September and Otis in mid-year. BSCS taught in 13 classes by
B 6 teachers -- Blue by one, Yellow by two and Green by three. Four teachers taught conventional biology* in 6 classes. All students took Watson-Glaser as post-test in May.

All students took DAT at start of trial year. Some groups* took Davis (2D) and others* took Illinois as pre-tests. Blue version studied by 3847, Green by 2500 and Yellow by
B 3499 students; who took different Quarterly Tests (R or S) and also BSCS Comprehensive Final (both I and II) at end of trial year. Some groups took Davis (2A) and others the Illinois as post-tests.

*No other information on this item.

	11. SAFEGUARDS TAKEN	12. OBSERVATIONS RECORDED
A	Evaluation Committee, aided by three sub-committees and Educational Testing Service, formulated evaluation program. Wide geographic, socio-economic and cultural population coverage. Continuous feedback in written comments, consultant visits, etc. (many examples cited).	No description of data collection methods. No examples of raw scores. ε
ε	None indicated.	Same as above. ε
B	Pre-test and IQ scores held constant. Pooling of data defended. Both t-test and F-values used. "The statistical techniques employed warrant confidence in the results obtained." (but see Welch, 9)	Same as above. ε
ε	None indicated.	S ame as above. ε

13. FINDINGS ASSEMBLED	14. STATISTICS INTERPRETED	
Data shown in 9 tables: Mean scores on 3 post-tests of achievement by: sex, treatment; sex and grade for non-block groups. Mean differences between various combinations. Correlations between ability and achievement scores. No data from attitude measures. Some verbal generalizations from questionnaires.	Mean scores, but no Standard Deviations. Correlations computed and levels of significance adduced (r's of -.09, .07, .17 and .24 called significant).	C
No tables of data, no numerical values reported. (Many tables in unpublished dissertation.)	F-values from four analyses of covariance used to accept or reject null hypotheses, at .05 level. Also some t-tests.	C
Data shown in 6 tables: Numbers of pupils, teachers and classes in groups. Summary of pooled data. Analysis of variance and covariance, adjusted means.	Adjusted means, t-tests, analysis of variance. Covariance to make pre-tests and IQ scores equivalent. Data pooled.	B
Data shown in 4 tables: Raw means, adjusted means and SD's for DAT, Quarterly and Comprehensive tests by: sex, BSCS version, test forms and by t-tests. Correlations between: Davis vs DAT, Davis vs Comprehensive, Illinois vs DAT, Illinois vs Comprehensive.	Product-moment correlations. Some t-tests.	D
A		

15. INTERPRETATIONS DISCUSSED

Correlations of SCAT scores vs end-of-year tests were high and "make it quite clear that performance on these three tests is highly dependent on academic ability." Differences among groups using three BSCS versions "were for the most part negligible and inconsistent from one test to the next." Males did better than females, yet "it is inappropriate to conclude that the boys were superior." "Superior 9th grade students do as well as unselected 10th graders." "The findings of the BSCS Evaluation Program -- through feedback and testing -- indicate" 9 listed generalizations (but some seem hardly warranted by data shown).

B

"The results of the study may indicate that students through the BSCS program learn the important core of information of the traditional plus the new, updated biology knowledge incorporated in the BSCS course." (Unwarranted by the data shown, and not even tested.)

D

Instruction in BSCS was neither inferior nor superior to conventional biology in improving critical thinking ability, within the terms and limitations of this study. (But no evidence that gains were due to biology teaching and not to increasing maturity, for no controls were used. Also see 9).

B

"None of the differences between pairs of test forms (R and S, I and II) is of any practical significance." "These differences (between three BSCS versions) demonstrate merely that one group does slightly better than another on these particular tests." "No consistent trends appear indicating that academic ability is a better predictor of one test than another." "Little, if any, gain would result from using both the DAT and a reading test in predicting BSCS final achievement."

A

16. CONCLUSIONS REACHED

"All BSCS groups substantially outperformed the control group on the BSCS Comprehensive Final." The control group "greatly excelled all BSCS groups on the Cooperative Biology Test" and "was slightly superior to all BSCS groups on the BSCS Impact
B Test." "While 11 of the 12 block-nonblock comparisons were significant, all differences were small." On two attitude measures, differences between BSCS and control groups were negligible. There was little or no relationship between Comprehensive Final results and traits of teachers or school.

"The experimental and control groups did not differ significantly on Nelson Biology Test." "The middle and high experimental groups excelled significantly over the middle and high control
C groups on BSCS Achievement test." "No significant differences appeared in achievement between middle and high ability levels on either of the two tests" nor "between the low ability groups on the BSCS test." "No significant differences appeared among schools on the Nelson but did so on the BSCS test."

Hypotheses 2 and 3 were accepted: Those in Green and Yellow versions were not significantly better than those in conventional
A biology. Hypotheses 1 and 4 were rejected: Pupils in the Blue version were significantly better than those in conventional biology and those in the Yellow version, who were in turn better than those in the Green version.

"BSCS students on the DAT were between 65th and 75th percentiles" of a national sample. "DAT was substantially related" to Quarterly and Comprehensive Final test performance. "Students
B in Blue version had the highest means on every ability, achievement and final test, while those in Green version had the lowest." "In most groups the males outperformed the females on all tests" but only by a few points. DAT was highly correlated with Davis and Illinois reading tests.

17. LIMITATIONS RECOGNIZED	18. FURTHER WORK PROJECTED
D Volunteer teachers and schools. Students above average ability. Some BSCS objectives were not susceptible to quantitative measurement.	Mentioned need for replication of evaluation studies. D
C Tests did not really measure low ability students. Teachers and and teaching climate may have influenced results. Population limited to a single community.	Check adequacy of tests for slow learners, and influence of teachers and teaching climate. Follow-up retention study of same pupils. Replicate with a more universal population. B
A Impossible to hold constant the competence, methods, experience, philosophy and preparation of teachers. With only one teacher for Blue version, high scores may have reflected his good teaching. Conclusions depended upon validity and reliability of instruments.	None proposed. E
E None acknowledged.	None proposed. E

19. IMPROVEMENTS SUGGESTED	20. CLARITY OF REPORT	
Clearly stated problem, objective and hypotheses. Fully described population and proper sampling. Instruments defended. Pre-tests. Assumptions and limitations recognized. Only warranted conclusions drawn. Better organization of report.	Writing clear and straightforward, with little jargon. Organization of report very poor, with chief elements scattered and buried. So much data deserved better planning, more rigor and more adequate reporting.	D
Larger population, better described. Procedure in full. Pre-tests with defended instruments. Tables with data and statistical treatment. Assumptions, safeguards, limitations recognized. Conclusions defended.	Written honestly; well organized and clear. But description of procedure and presentation of results much too scanty, leaving conclusions suspect.	B
Larger population and proper sampling. More carefully chosen teachers. Standardized teaching program. Replication within project.	Relatively little jargon, but considerable repetition. Clear descriptions and good organization.	B
Clearly stated problem, objective and hypotheses. Proper sampling of fully described population. Pre-tests on achievement. Control groups. Assumptions, safeguards, limitations recognized. Better statistical treatment of data. Conclusions fully defended.	Writing good, with little jargon. But a report for general readers should have given more information (even if in BSCS Manual) on students, procedure, and instruments.	C

REFERENCES

1. ERIC Information Analysis Center for Science Education, 1460 West Lane avenue, Columbus, Ohio 43210.
2. Lance, Mary L. "A Comparison of Gains in Achievement Made by Students of BSCS High School Biology and Students of a Conventional Course in Biology." Unpublished doctoral dissertation, Univ. of Georgia. Diss. Abstr. 25:2814, 1964.
3. Lewis, William A. "An Evaluation of Four Selected Approaches to Teaching High School Biology." Unpublished doctoral dissertation, Montana State Univ. Diss. Abstr. 27A:1689, 1966.
4. Moore, C. Olan. "An Evaluation of the Effectiveness of the Biological Sciences Curriculum Study Approach to Teaching Biology to High Ability Students in the Ninth Grade." Unpublished doctoral dissertation, Arizona State Univ. Diss. Abstr. 26A:3129, 1965.
5. Rozolis, James T. "A Comparison of Traditional High School Biology Courses and the Biological Sciences Curriculum Study Program of Courses." BioScience, 25:315-324, May 1967.
6. Schuck, Robert F. "The Effects of Set Induction Upon the Achievement of Ninth-Grade Pupils and Their Perception of Teacher Effectiveness." Journal of Educational Research, 62:279-285, February 1969.
7. Strauss, Samuel. "Guidelines for Analysis of Research Reports." Journal of Educational Research, 63:165-169, December 1969.
8. Wandt, Edwin. An Evaluation of Educational Research Published in Journals. Report of Committee on Evaluation of Research. Washington: American Educational Research Ass'n, Feb. 1967.
9. Welch, Wayne W. "Curriculum Evaluation." Review of Educational Research, 39:429-443, October 1969.
10. Welch, W.W., Walberg, H.J. and Ahlgren, A. "The Selection of a National Random Sample of Teachers for Experimental Curriculum Evaluation." School Science and Mathematics, 69:210-216, March 1969.