DOCUMENT RESUME

ED 267 C88 SE 046 510

AUTHOR	Shymansky, James A.; And Others
TITLE	A Study of Uncertainties in the Meta-Analysis of Research on the Effectiveness of "New" Science Curricula. Prelim'nary Report.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	Mar 86
GRANT	NSF-MDR-8550470
NOTE	28p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (59th, San Francisco, CA, March 28 - April 1, 1986).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Achievement; Elementary Secondary Education; *Meta Analysis; *Program Effectiveness; *Research Methodology; *Science Course Improvement Projects; *Science Education; Science Programs
IDENTIFIERS	National Science Foundation; Science Education Research

ABSTRACT

In 1981 a meta-analysis of research in science education was completed by a team of researchers under the direction of Ronald Anderson at the University of Colorado. Individual teams synthesized the extant research using meta-analysis procedures. One of the syntheses showed that the post-Sputnik or "new" science programs on the whole produced very positive gains in student achievement. Statisticians have questioned the integrity of meta-analysis results because standard procedures fail to take into account the error associated with the sample estimation of the effect size. It is argued that when a set of different effect sizes are pooled across studies, the means most properly should be computed using weighted factors associated with the precision of the effect size values. This research repeated the meta-analysis of research on the new science curricula using weighted procedures proposed by Hedges. The results provide more precise information on the actual effectiveness of the post-Sputnik science curricula meta-analyzed in the Ander[illegible]ject and establish clearer guidelines regarding the application [illegible]eta-analysis techniques in synthesizing research results. (A[illegible]/JN)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED267988

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

PRELIMINARY REPORT

A STUDY OF UNCERTAINTIES IN THE META-ANALYSIS OF RESEARCH ON THE EFFECTIVENESS OF "NEW" SCIENCE CURRICULA*

James A. Shymansky
George Woodworth
Craig Berg
University of Iowa

Larry V. Hedges
University of Georgia

ANNUAL MEETING
National Association for Research in Science Teaching
Meridien Hotel
San Francisco
March 29, 1986

*This research was supported by the National Science Foundation under Grant No. MDR-8550470. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
James A. Shymansky

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE046510

ERIC

Preliminary Report

A STUDY OF UNCERTAINTIES IN THE META-ANALYSIS OF RESEARCH ON THE EFFECTIVENESS OF "NEW" SCIENCE CURRICULA

ABSTRACT

In 1981 a meta-analysis of research in science education was completed by a team of researchers under the direction of Ronald Anderson at the University of Colorado (NSF Grant SED 80-12310). Individual teams synthesized the extant research using "meta-analysis" procedures developed by Glass (1976). One of the syntheses showed that the post-Sputnik or "new" science programs on the whole produced very positive gains in student performance (Shymansky, Kyle, and Alport, 1983).

Statisticians have questioned the integrity of meta-analysis results because standard procedures fail to take into account the error associated with the sample estimation of the effect size. It is argued that when a set of different effect size are pooled across studies, the means most properly should be computed using weighting factors associated with the precision of the effect size values (Hedges, 1982, 1982b).

This research project repeated the meta-analysis of research on the new science curricula using weighting procedures proposed by Hedges. The results of this study provide more precise information on the actual effectiveness of the post-Sputnik science curricula meta-analyzed in the Anderson project and establish clearer guidelines regarding the application of meta-analysis techniques in synthesizing research results.

A STUDY OF UNCERTAINTIES IN THE META-ANALYSIS OF RESEARCH ON THE EFFECTIVENESS OF "NEW" SCIENCE CURRICULA

PURPOSE

The purpose of this project was to verify the results of the meta-analysis of the research on the effectiveness of post-Sputnik science curricula (Shymansky, Kyle, and Alport, 1983) using procedures proposed by Hedges (1981, 1982, 1982b). The reanalysis took into account the standard error inherent in the sample-based effect size estimates obtained using the procedures outlined by Glass (1976, 1984). Mean effect sizes determined in the Shymansky et al synthesis were recalculated using weighted values of effect sizes.

BACKGROUND

The 1981 meta-analysis of research on the effectiveness of post-Sputnik science curricula (Shymansky, et al, 1983) quantitatively synthesized the results of 105 studies. The research was part of a comprehensive meta-analysis of research in science education under the direction of Ronald Anderson at the University of Colorado (NSF Grant SED 80-12310). Gene Glass and Mary Lee Smith, originators of the meta-analysis procedure, served as consultants to the project. The synthesis showed that the "new" science curricula uniformly increased student performance on affective and cognitive criteria by as much as 0.34 standard deviations.

The question of precision of the effect size estimates obtained from the aggregated studies was handled by testing for differences in mean effect size between sub-groups of aggregates based on such a priori factors as sample size, length of treatment, judged internal validity, and form of publication. In all these sub-group analyses, no significant differences were found. The no-difference results were viewed as grounds for giving equal weight to all effect sizes in the aggregation procedures.

In the four or so years since the Anderson meta-analysis project was conducted, statisticians have examined the meta-analysis procedures carefully and have questioned some of the techniques and underlying assumptions (Slavin, 1984; Eysenck, 1978). One area receiving considerable attention deals with the precision of the calculated effect size. Assuming a series of k independent studies share a common effect size δ, it is argued that it is reasonable to give more weight to the more precise estimates of when pooling study results (Hedges, 1982).

According to Glass (1976), an estimate of effect size d is defined by

$$(1)\quad d = \frac{Y^{E} - Y^{C}}{S}$$

where Y^{E} and Y^{C} are the experimental and control group sample means and S is the pooled within-group sample standard deviation. An examination of the effect size estimate d reveals that it is a slightly biased estimator of δ that tends to overestimate δ for small samples. A simple correction gives an unbiased estimator of δ (Hedges, 1981). This unbiased estimator, d is obtained by multiplying d by a constant that depends on the sample size in the study.

$$(2) \quad d = c_n d = c_n(\bar{Y}^E - \bar{Y}^C)/S$$

where the values of c_n are given to a good approximation by

$$(3) \quad c_n = 1 - \frac{3}{4n^E + 4n^C - 9}$$

Hedges (1981) showed that d is approximately normally distributed with mean δ variance

$$(4) \quad v = \frac{n^E + n^C}{n^E n^C} + \frac{d^2}{2(n^E + n^C)}$$

The variance of d is completely determined by the sample sizes and the value of d. Consequently, it is possible to determine the sampling variance of d from a single observation of d is the key to modern statistical methods for meta-analysis. This relationship allows the meta-analyst to use all the degrees of freedom among the different d values for estimating systematic effects while still providing a way of estimating the unsystematic variance needed to construct statistical tests (Hedges, 1984).

Of primary importance to the meta-analyst is the question of how to combine the estimates of a series of effect sizes. One way of combining the estimates is simply to take the average d. The most precise combination, however, is weighted average that takes the variances $v_1, \ldots v_k$ of the effect size estimates $d_1, \ldots d_k$ into account.

ERIC

Hedges (1981, 1982, 1982a) outlines a procedure for obtaining a weighted estimate of effect size δ.

Once a weighted estimate of D is known, tests for homogeneity of effect size can be made before they are pooled and confidence intervals for mean effect sizes can be calculated (Hedges, 1982). If all k studies share a common effect size δ, the weighted mean D is approximately normally distributed with a mean of δ and a variance of

$$(5)\quad v^* = 1 \Big/ \sum_{i=1}^{k} \frac{2(n_i^E + n_i^C)n_i^E n_i^C}{2(n_i^E + n_i^C)^2 + n_i^E n_i^C d_i^2}$$

Consequently, if it is reasonable to believe that a set of studies shares a common effect size δ, then a 100 (1 - α)

$$(6)\quad D - z_\alpha \sqrt{v^*} < \delta < D + z_\alpha \sqrt{v^*}$$

where z_α is the 100 α percent two-tailed value of the standard normal distribution. If the confidence interval does not include zero, then the hypothesis that $\delta = 0$ is rejected at the significance level α (Hedges, 1984).

PROCEDURES

The raw data from the Shymansky et al study of the effectiveness of new science curricula were obtained from computer records for re-analysis. Original research reports were retrieved and examined for critical data not found in the existing computer records.

Computer programs were written to calculate the weighted effect size estimates using the formulas provided by Hedges (1982, 1984). Weighted effect sizes were aggregated as in the original Shymansky et al meta-analysis. Means obtained from weighted and unweighted effect sizes were compared to determine the influence of the weighting procedure on the conclusions of the original project.

In addition, tests for homogeneity of effect size were conducted for the aggregate groups studied by Shymansky, et al using procedures outlined by Hedges (1982, 1982a). In cases where effect sizes were found to be inhomogeneous across the study groups, explanations for the variations in effect sizes in terms of study characteristics were explored.

RESULTS

Comparisons of mean effect sizes for selected sub-groups using unweighted effect size estimates from the original 1981 meta-analysis and weighted effect size estimates adjusted for sample size, multiple measures, and covariates are presented in Tables 1-13. While these tables do not represent the complete complement of analyses performed, they illustrate the major differences resulting from the use of the modern procedures.

DISCUSSION

Several interesting differences are evident in the comparison of the 1981 and 1986 analyses. Only 81 of the 105 studies included in the 1981 analysis survived in the re-analysis. Reasons for dropping the 24 studies included lack of data and inappropriate unit of analysis.

The procedures for aggregating effect sizes represent another major difference in the 1981 and 1986 analyses. In the 1981 project all effect size estimates were treated as independent values in the aggregations. In the 1986 study effect size estimates were averaged within sub-groups before aggregation. Thus, at the aggregate level, the effect size estimates included represent independent subgroups of students. For example, in Table 1, the "composite" effect size mean is based on 136 independent sub-groups based on the averages of 321 effect size estimates whereas the "criterion clusters" (achievement, perceptions, etc.) effect size means are based on separate independent sub-groups of 84, 18, 47, 25, 17, 9 (for a total of 200). Within any one criterion cluster, the sub-groups are independent, but between clusters they are not.

It is interesting to note that the changes in mean effect sizes between the 1981 and 1986 analyses are not uniform. In some cases the means decrease (Table 1), in others they increase or don't change at all. The important factor is that the means calculated using the refined procedures are more precise, more valid than those generated in the 1981 study. Like any piece of quality research, the 1986 results carefully account for sources of systematic error in the data source.

ERIC

TABLE 1

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ALL STUDIES X ALL STUDENTS

	STANDARD (1981)		REFINED (1986)		
CRITERION	N	DBAR	N	DBAR	95% CI
COMPOSITE	340	0.34*	136	0.26*	±0.08
ACHIEVEMENT	130	0.37*	84	0.30*	±0.12
PERCEPTIONS	51	0.37*	18	0.19*	±0.12
PROCESS	56	0.39*	47	0.33*	±0.12
ANALYTIC	35	0.25*	25	0.13	±0.16
RELATED	46	0.25*	17	-0.10	±0.15
OTHER	21	0.33*	9	0.10	±0.39

*Significant at the $\alpha = 0.05$ level

TABLE 2

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ALL STUDIES X GENDER

	STANDARD (1981)		REFINED (1986)		
CRITERION	N	DBAR	N	DBAR	95% CI
MALE SS					
COMPOSITE	123	0.22*	59	0.24*	±0.12
ACHIEVEMENT	58	0.25*	39	0.28	±0.17
PERCEPTIONS	12	-0.02	4	0.19*	±0.18
PROCESS	18	0.16	12	0.27*	±0.17
ANALYTIC	21	0.30*	15	0.19	±0.22
FEMALE SS					
COMPOSITE	19	0.25*	7	0.56*	±0.54
ACHIEVEMENT	4	0.55	5	0.69	±1.28
PERCEPTIONS	5	0.32	1	0.54	±0.56
PROCESS	5	0.29*	3	0.32	±0.56
ANALYTIC	5	-0.10	1	0.35	±0.21
MIXED SS					
COMPOSITE	199	0.43*	70	0.26*	±0.12
ACHIEVEMENT	68	0.45*	40	0.29*	±0.16
PERCEPTIONS	34	0.51*	13	0.18*	±0.14
PROCESS	33	0.52*	32	0.36*	±0.16
ANALYTIC	9	0.31	9	0.06	±0.27

*Significant at the $\alpha = 0.05$ level

TABLE 3

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ALL STUDIES X SES

	STANDARD (1981)		REFINED (1986)		
CRITERION	N	DBAR	N	DBAR	95% CI
LOW SES					
COMPOSITE	4	0.63*	2	0.28	±1.14
ACHIEVEMENT	1	1.08	1	0.78	±0.84
PROCESS	1	0.64	1	0.13	±0.40
RELATED	2	0.41	-	-	-
MID SES					
COMPOSITE	302	0.28*	127	0.27*	±0.08
ACHIEVEMENT	105	0.27*	77	0.31*	±0.12
PERCEPTIONS	49	0.32*	16	0.19*	±0.12
PROCESS	49	0.33*	42	0.36*	±0.12
ANALYTIC	33	0.23*	21	0.14	±0.18
RELATED	46	0.24*	16	-0.07	±0.18
OTHER	20	0.31*	9	0.10	±0.38
HIGH SES					
COMPOSITE	19	0.99*	7	0.03	±0.27
ACHIEVEMENT	11	1.00*	6	-0.04	±0.45
PERCEPTIONS	2	1.40	2	-0.01	±0.26
PROCESS	4	1.00	4	0.04	±0.37
ANALYTIC	2	0.50	4	-0.10	±0.18

*Significant at the α = 0.05 level

TABLE 4

COMPARISON OF EFFECT DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ALL STUDIES X GRADE LEVEL

CRITERION	STANDARD (1981) N	DBAR	REFINED (1986) N	DBAR	95% CI
ELEMENTARY (K-6)					
COMPOSITE	124	0.31*	54	0.23*	±0.14
ACHIEVEMENT	27	0.37*	23	0.15	±0.24
PERCEPTIONS	29	0.28*	6	0.45*	±0.18
PROCESS	16	0.56*	19	0.51*	±0.24
ANALYTIC	1	0.06	2	0.19	±0.35
RELATED	37	0.17*	14	-0.02	±0.20
OTHER	14	0.32*	8	0.18	±0.39
JUNIOR HIGH (7-9)					
COMPOSITE	72	0.31*	22	0.33*	±0.13
ACHIEVEMENT	13	0.23*	10	0.39*	±0.25
PERCEPTIONS	11	0.59*	3	0.33	±0.44
PROCESS	18	0.23	12	0.39*	±0.13
ANALYTIC	14	0.02	5	0.23*	±0.18
RELATED	9	0.68	6	0.10	±0.33
OTHER	7	0.33	-	-	-
HIGH SCHOOL (10-12)					
COMPOSITE	132	0.38*	64	0.25*	±0.12
ACHIEVEMENT	83	0.37*	51	0.30*	±0.14
PERCEPTIONS	9	0.44	5	0.11*	±0.08
PROCESS	19	0.43*	18	0.21*	±0.16
ANALYTIC	19	0.42*	17	0.08	±0.25
RELATED	2	-0.23	3	-0.15*	±0.10

*Significant at the α = 0.05 level

ERIC

TABLE 5

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ALL STUDIES X SCHOOL TYPE

CRITERION	STANDARD (1981) N	STANDARD (1981) DBAR	REFINED (1986) N	REFINED (1986) DBAR	REFINED (1986) 95% CI
RURAL					
COMPOSITE	25	0.20*	12	0.22*	±0.13
ACHIEVEMENT	9	0.34*	11	0.21	±0.30
PERCEPTIONS	9	-0.07	1	0.48	±0.38
PROCESS	6	0.45*	8	0.35*	±0.25
SUBURBAN					
COMPOSITE	168	0.38*	70	0.22*	+0.12
ACHIEVEMENT	72	0.41*	40	0.25*	+0.18
PERCEPTIONS	19	0.46*	12	0.12*	+0.06
PROCESS	13	0.50*	14	0.30*	+0.22
ANALYTIC	17	0.27*	17	0.14	+0.16
RELATED	34	0.30*	13	-0.09	+0.20
URBAN					
COMPOSITE	32	0.34*	8	0.68*	±0.25
ACHIEVEMENT	4	0.81*	5	0.81*	±0.22
PERCEPTIONS	2	0.64*	-	-	-
PROCESS	12	0.24	3	0.31*	±0.13
ANALYTIC	11	0.17	2	0.40*	±0.14
RELATED	2	0.41	-	-	-

*Significant at the $\alpha = 0.05$ level

TABLE 6

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ALL STUDIES X K-8 SUBJECT AREAS

CRITERION	STANDARD (1981) N	DBAR	REFINED (1986) N	DBAR	95% CI
<u>GENERAL SCIENCE</u>					
COMPOSITE	128	0.35*	54	0.23*	±0.14
ACHIEVEMENT	32	0.35*	23	0.15	±0.24
PERCEPTIONS	30	0.32*	6	0.45	-
PROCESS	19	0.59*	19	0.51*	±0.24
ANALYTIC	1	0.06	2	0.19	±0.35
RELATED	46	0.27*	14	-0.02	±0.20
<u>LIFE SCIENCE</u>					
COMPOSITE	5	0.53*	2	0.36	±0.60
PERCEPTIONS	4	0.66	1	0.63	±0.21
ANALYTIC	1	0.01	1	0.03	±0.22
<u>PHYSICAL SCIENCE</u>					
COMPOSITE	40	0.17*	12	0.33*	±0.11
ACHIEVEMENT	9	0.31	7	0.47*	±0.27
PERCEPTIONS	8	0.31*	6	0.11	-
PROCESS	10	0.08	9	0.34*	±0.19
ANALYTIC	7	-0.10	4	0.26*	±0.19
<u>EARTH SCIENCE</u>					
COMPOSITE	20	0.14	2	0.07	±0.19
ACHIEVEMENT	4	-0.07	1	-0.04	±0.14
PERCEPTIONS	1	0.11	1	0.11	±0.14
PROCESS	8	0.22	1	0.21	±0.24
ANALYTIC	7	0.16	1	0.15	±0.14

*Significant at the $\alpha = 0.05$ level

TABLE 7

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ALL STUDIES X HIGH SCHOOL SUBJECTS

CRITERION	STANDARD (1981) N	DBAR	REFINED (1986) N	DBAR	95% CI
BIOLOGY					
COMPOSITE	47	0.61	29	0.33*	±0.18
ACHIEVEMENT	29	0.59*	24	0.43*	±0.17
PERCEPTIONS	4	0.82	3	0.11	±0.12
PROCESS	6	0.90	7	0.22	±0.23
ANALYTIC	7	0.46	7	-0.05	±0.60
RELATED	1	-0.50	1	-0.17	±0.12
CHEMISTRY					
COMPOSITE	49	0.16	19	0.10	±0.22
ACHIEVEMENT	33	0.16*	17	0.13	±0.25
PERCEPTIONS	4	0.15	1	0.15	±0.57
PROCESS	6	0.02	6	0.13	±0.40
ANALYTIC	6	0.28	7	0.26	±0.34
PHYSICS					
COMPOSITE	37	0.46	18	0.28*	±0.22
ACHIEVEMENT	23	0.50*	12	0.35	±0.37
PROCESS	7	0.33	5	0.31*	±0.20
ANALYTIC	6	0.53	3	-0.03	±0.16
RELATED	1	0.04	1	-0.02	±0.17

*Significant at the $\alpha = 0.05$ level

ERIC

TABLE 8

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ALL STUDIES X HIGH SCHOOL SUBJECTS

	STANDARD (1981)		RANDOMIZED/MATCHED DESIGNS (1986)		
CRITERION	N	DBAR	N	DBAR	95% CI
<u>BIOLOGY</u>					
COMPOSITE	47	0.61*	10	0.30*	±0.26
ACHIEVEMENT	29	0.59*	10	0.40*	±0.33
PERCEPTIONS	4	0.82	2	-0.01	±0.27
PROCESS	6	0.90	6	0.16	±0.32
ANALYTIC	7	0.46	2	-0.11	±0.26
RELATED	1	-0.50	-	-	-
<u>CHEMISTRY</u>					
COMPOSITE	49	0.16*	4	0.31*	±0.18
ACHIEVEMENT	33	0.16*	4	0.31*	±0.14
PERCEPTIONS	4	0.15	-	-	-
PROCESS	6	0.02	2	0.25*	±0.18
ANALYTIC	6	0.28	2	0.32	±0.32
<u>PHYSICS</u>					
COMPOSITE	37	0.46*	2	0.26*	±0.12
ACHIEVEMENT	23	0.50*	1	0.26	±0.14
PROCESS	7	0.33	1	0.24	±0.19
ANALYTIC	6	0.53	-	-	-
RELATED	1	0.04	-	-	-

*Significant at the $\alpha = 0.05$ level

TABLE 9

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ALL STUDIES X HIGH SCHOOL SUBJECTS

	STANDARD (1981)		STANDARDIZED TESTS ONLY (1986)		
CRITERION	N	DBAR	N	DBAR	95% CI
BIOLOGY					
COMPOSITE	47	0.61	20	0.18	±0.20
ACHIEVEMENT	29	0.59*	15	0.28*	±0.22
PERCEPTIONS	4	0.82	3	0.11	±0.12
PROCESS	6	0.90	7	0.22	±0.23
ANALYTIC	7	0.46	7	-0.05	±0.60
RELATED	1	-0.50	-	-	-
CHEMISTRY					
COMPOSITE	49	0.16	10	0.15	±0.27
ACHIEVEMENT	33	0.16*	8	0.10	±0.33
PERCEPTIONS	4	0.15	-	-	-
PROCESS	6	0.02	4	0.07	±0.21
ANALYTIC	6	0.28	4	0.35	±0.38
PHYSICS					
COMPOSITE	37	0.46	14	0.26	±0.28
ACHIEVEMENT	23	0.50*	8	0.33	±0.57
PROCESS	7	0.33	4	0.39*	±0.14
ANALYTIC	6	0.53	3	-0.03	±0.16
RELATED	1	0.04	1	-0.02	±0.17

*Significant at the $\alpha = 0.05$ level

TABLE 10

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ACHIEVEMENT X CURRICULUM

CURRICULUM	STANDARD (1981) N	DBAR	REFINED (1986) N	DBAR	95% CI
ESS	3	0.09	4	0.04	±0.36
SCIS	5	1.00	2	1.09*	±0.28
SAPA	12	0.17	11	0.03	±0.26
MNMT	2	1.51	1	1.72	±0.32
ESTPSI	3	0.28	4	0.26*	±0.18
FHESP	1	-0.06	3	0.24	±0.53
IPS	3	0.03	2	0.28	±0.49
ESCP	6	0.19	2	0.17	±0.85
IME	2	-0.11	1	0.20	±0.25
MSP	1	0.42	1	0.49*	±0.12
BSC-S	2	0.02	2	0.44	±0.54
BSC-Y	19	0.45*	18	0.47*	±0.19
BSC-B	2	3.94*	1	1.01	±0.53
BSC-G	2	0.17	1	0.01	±0.17
BSC-A	4	0.09	2	0.11	±0.26
CHEMS	23	0.12	13	0.03	±0.24
CBA	10	0.27	4	0.53*	±0.45
PSSC	23	0.51	12	0.34	±0.37

*Significant at the $\alpha = 0.05$ level

TABLE 11

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ACHIEVEMENT X CURRICULUM

	STANDARD (1981)		RANDOMIZED/MATCHED DESIGNS (1986)		
CRITERION	N	DBAR	N	DBAR	95% CI
ESS	3	0.09	1	0.04	±0.38
SCIS	5	1.00	1	1.12	±0.29
SAPA	12	0.17	5	0.14	±0.60
MNMT	2	1.51	-	-	-
ESTPSI	3	0.28	-	-	-
FHESP	1	-0.06	-	-	-
IPS	3	0.03	1	0.56	±0.32
ESCP	6	0.19	1	0.74	±0.22
IME	2	-0.11	-	-	-
MSP	1	0.42	-	-	-
BSC-S	2	0.02	-	-	-
BSC-Y	19	0.45*	7	0.47*	±0.43
BSC-B	2	3.94*	1	1.01	±0.53
BSC-G	2	0.17	-	-	-
BSC-A	4	0.09	2	0.11	±0.26
CHEMS	23	0.12	3	0.32*	±0.23
CBA	10	0.27	1	0.29	±0.18
PSSC	23	0.51	1	0.27	±0.14

*Significant at the α =0.05 level

TABLE 12

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

ACHIEVEMENT X CURRICULUM

CRITERION	STANDARD (1981) N	DBAR	STANDARDIZED TESTS ONLY (1986) N	DBAR	95% CI
ESS	3	0.09	4	0.04	±0.36
SCIS	5	1.00	-	-	-
SAPA	12	0.27	11	0.03	±0.27
MNMT	2	1.51	-	-	-
ESTPSI	3	0.28	-	-	-
FHESP	1	-0.06	3	0.24	±0.53
IPS	3	0.03	2	0.28	±0.48
ESCP	6	0.19	1	0.74	±0.24
IME	2	-0.11	1	0.20	±0.25
MSP	1	0.42	1	0.49	±0.12
BSC-S	2	0.02	2	0.44	±0.54
BSC-Y	19	0.45*	9	0.33*	±0.32
BSC-B	2	3.94*	1	1.01	±0.53
BSC-G	2	0.17	1	0.01	±0.17
BSC-A	4	0.09	2	0.11	±0.26
CHEMS	23	0.12	5	0.01	±0.46
CBA	10	0.27	3	0.32*	±0.16
PSSC	23	0.51	8	0.33	±0.57

*Significant at the $\alpha = 0.05$ level

TABLE 13

COMPARISON OF EFFECT SIZE DATA OBTAINED USING STANDARD AND REFINED ESTIMATION PROCEDURES

* * *

COMPOSITE X INSERVICE

CRITERION	STANDARD (1981) N	STANDARD (1981) DBAR	REFINED (1986) N	REFINED (1986) DBAR	REFINED (1986) 95% CI
INSERVICE	112	0.23*	49	0.27*	±0.11
NO INSERVICE	14	0.50*	5	0.23	±0.30
NOT REPORTED	215	0.38*	82	0.25*	±0.12

*Significant at the α = 0.05 level

REFERENCES

Eysenck, H. J., "An exercise in mega-silliness." American Psychology, 1978, 33, 517.

Glass, G. V. "Primary, secondary, and meta-analysis of research." Educational Researcher, 1976, 5, 3-8.

Glass, G. V., McGaw, B. and Smith, M. L. Meta-Analysis in Social Research. Beverly Hills: Sage, 1984.

Hedges, L. V. "Distribution theory for Glass's estimator of effect size and related measures." Journal of Educational Statistics, 1981 6, 107-128.

Hedges, L. V. "Estimation of effect size from a series of independent experiments." Psychological Bulletin, 1982, 92, 490-499.

Hedges, L. V. "Fitting categorical models to effect sizes from a series of experiments." Journal of Educational Statistics, 1982a, 7, 119-137.

Hedges, L. V. "Fitting continuous models to effect size data." Journal of Educational Statistics, 1982b, 7, 245-270.

Hedges, L. V. "Advances in statistical methods for meta-analysis." In W. H. Yeaton, P. M. Wortman (Eds.), Issues in Data Synthesis. San Francisco: Josey-Bass, December, 1984, 25-42.

Shymansky, J. A., Kyle, W. C., and Alport, J. M. "The effects of new science curricula on student performance." Journal of Research in Science Teaching, 1983, 20, 387-404.

Slavin, R. E. "Meta-analysis in education: How has it been used?" Educational Researcher, 1984, October, 6-27.

APPENDIX

APPENDIX

REVISED META-ANALYSIS CODING SHEET
(September 13, 1985)

COLUMNS	DESCRIPTOR
1-3	Study Number
4-5	% Non-White (indicates any minority)
6-7	Study Length (in weeks)
8	School Size (1) < 50 (2) 50-199 (3) 200-499 (4) 500-999 (5) 1000-1999 (6) > 2000
9	School Type (1) Rural (2) Suburban (3) Urban
10-11	Publication date (Year)
12	Publication form (1) Journal (2) Book (3) MA/MS Thesis (4) Dissertation (5) Unpublished
13	Curriculum Profile-Inquiry (Rating 1 (low) - 4 (High)
14	Curriculum Profile-Process Skills (1-4)
15	Curriculum Profile-Laboratory Emphasis (1-4)
16	Curriculum Profile-Individualization (1-4)
17	Curriculum Profile - Content Emphasis (1-4)
18-19	% Female Teachers
20-21	years Experience - Treatment Teachers
22-23	Years Experience - Control Teachers
24-25	%Non-White Teachers in Treatment Group
26	Educational Background of Treatment Teachers (1) No bachelors (2) BS (3) BS + 30 (4) MS (5) MS + 15 (6) MS + 30 (7) Doctorate
27	Funding of Inservice (1) Locally (2) University (3) Federally (4) Unknown
28	Subject assignment (1) Random (2) Matched (3) Intact (4) Self-selected
29	Unit of Analysis (1) Individual (2) Classroom (3) School (4) Other Group

30 Type of Study (1) Correlational (2) Quasi-Experimental (3) Experimental

31 Internal Validity (1) Low (intact, dissimilar) (2) Medium (3) High

32 Reported Significance (1) .005 (2) .01 (3) .05 (4) .10 (5) $>$.10

33-34 Effect Size Source (1) Directly from data (2) Estimated from report

35 Source of Means (1) Unadjusted post-test (2) Estimated from report

37 Method of Measurement (1) Standardized (2) Ad hoc written test (3) Class Test (4) Observation (5) Structured Interview

38 Preservice for Treatment Teachers

39 Inservice for Treatment Teachers

40 Content Measure (1) Life (2) Physical (3) General (4) Earth (5) Biology (6) Chemistry (7) Physics

41-42 Treatment

ELEMENTARY

01 ESS
02 SCIS, SCIIS, SCIS II
03 S-APA
04 OBIS
05 ESLI
06 ESSENCE
07 COPES
08 MAPS
09 USMES
10 MINNEMAST
11 IS
12 SCIL
15 ESTPSI
16 FHESP

JUNIOR HIGH CURRICULA

28 HSP
29 TSM
30 ISIS
31 ISCS
33 IPS
34 ESCP
35 IME
36 CE/EE
37 MSP

SECONDARY CURRICULA

50 BSCS (Special Materials)
51 BSCS (Yellow)
52 BSCS (Blue)
53 BSCS (Green)
54 BSCS (Advanced)
55 CHEM Study
56 CBA
57 PSSC
58 HPP
59 CE/EE
60 PSNS
61 IAC

43 Grade Level (1) K-3 (2) 4-6 (3) 7-9 (4) 10-12 (5) 12

44 Student Gender (1) >75% male (2) >75% female (3) Mixed

45 IQ (1) Low (2) Medium (3) High (Treat blanks, 4=2)

46 SES (1) Low (2) Medium (3) High (Treat blanks. 4=2)

47-48 Criterion Measure

1. Cognitive - low
2. Cognitive - high
3. Cognitive - mixed/general achievement
4. Problem solving
5. Affective - subject
6. Affective - science
7. Affective - procedure/methodology
8. Values**
9. Process skills, techniques
10. Methods of Science
11. Psychomotor**
12. Critical thinking
13. Creativity
14. Decision making**
15. Logical thinking (Piagetian)
16. Spatial relations (Piagetian)
17. Self-concept
18. Classroom behaviors (on tasks, etc.)**

19. Reading (comprehension, readiness)
20. Mathematics (concepts/skillsd, applications)
21. Social studies (content, skills)
22. Communication skills

49-50	Subject Subgroup (Subsamples (A, B, C, . . .) or repeated measures (A1, A2, . . .) within one study)
51-55	Mean - Treatment
56-60	STD Dev - Treatment
61-64	N - Treatment
65-69	Mean - Control
70-74	STD Dev - Control
75-78	N - Control
79-83	SS or MS Curriculum
84	Df Curriculum
85-92	SS Total
93-95	Df Total
96-100	F Curriculum
101-105	Effect Size
107	Special Flag