DOCUMENT RESUME

ED 270 328 SE 046 635

AUTHOR	Becker, Betsy Jane; Chang, Lin
TITLE	Measurement of Science Achievement and Its Role in Gender Differences.
INSTITUTION	Michigan State Univ., East Lansing.
PUB DATE	86
NOTE	30p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Data Interpretation; Females; Measurement Techniques; *Meta Analysis; Research Methodology; Science Education; Secondary Education; *Secondary School Science; *Sex Differences
IDENTIFIERS	Science Education Research

ABSTRACT

Differences in science achievement between males and females have been examined either directly or indirectly in a variety of studies. This investigation reviewed a quantitative synthesis of correlational research on science affect, ability, and achievement conducted by Steinkamp and Maehr. Their findings were reassessed by employing a meta-analysis approach which used tests for fitting categorical models to effect sizes. The reexamination focused on explanations of the reported differences in science achievement between males and females as well as on the role of measurement variables in the size of the gender differences. Results indicated that though gender differences tended to favor males, even significant differences were slight, and gender differences for many subsets of studies were not significant. The size of the gender difference depended in part on the science subject matter being tested and also on the type of measure used in the studies. A reference list and a list of synthesized studies are appended. (ML)

The Measurement of Science Achievement and its Role in Gender Differences

Betsy Jane Becker and Lin Chang

Michigan State University

Paper to be presented at the annual meeting of the American Educational Research Association. San Francisco, April, 1986.

The Measurement of Science Achievement and its Role in Gender Differences

Betsy Jane Becker and Lin Chang
Michigan State University

Abstract

It is often thought that women have not achieved in the area of science to the same degree as have men. Various studies have either directly or indirectly examined differences in science achievement between males and females, and have found contradictory results. This quantitative review or meta-analysis of studies of gender differences in science achievement addresses the question of whether males have higher achievement levels in science than females, and explores the role of measurement variables in the size of the gender differences.

We review a set of studies of sex differences in science achievement gathered from two earlier reviews by Steinkamp and Maehr. Steinkamp and Maehr conducted a quantitative synthesis of correlational research on science affect, ability, and achievement in which they reported conclusions based on average correlations. Later, they reviewed another set of studies (using effect sizes) and again reported only the average effect size across all studies.

The purpose of our study is to reexamine part of Steinkamp and Maehr's work by focusing on explanations for the differences in science achievement between males and females, using Hedges's approach to meta-analysis. Our review also improves upon the earlier reviews by applying Glass's effect sizes together with Hedges's tests for model fit to the studies of science-achievement gender differences collected by Steinkamp and Maehr.

Results indicated that though gender differences tended to favor males, even significant differences were slight, and gender differences for many subsets of studies were not significant. The size of the gender difference depended in part on the science subject matter being tested and also on the type of measure used in the studies.

Though the subject-matter content of the science-achievement measure was significantly related to the size of science gender differences, unexplained variation in the gender differences still remained after all of our measurement-related explanatory factors had been explored.

The Measurement of Science Achievement and its Role in Gender Differences

In the past decade issues of gender equity in education and the workplace have received increasing concern (e.g., Brickley, Garfunkel & Hulsizer, 1979). A long-standing finding which has received some attention is the dearth of women in scientific careers (e.g., National Science Foundation, 1977). This scarcity has been linked by some to earlier gender differences in science achievement (e.g., DeBoer, 1984).

Two recent reviews have summarized many studies which allow the examination of gender differences in science achievement (Steinkamp & Maehr, 1983, 1984). These reviews report results from several hundred samples which seem to indicate a general superiority of males on measures of science achievement, though considerable variation in the size and even the direction of the gender difference can be discerned across samples.

In this paper we attempt to understand more about gender differences in science achievement by reanalyzing the studies reviewed by Steinkamp and Maehr. By examining variation in the gender differences and the relationship of achievement differences to explanatory variables we hope to gain a better understanding of the interrelationship of gender and science achievement. We consider as possible explanatory variables features of the studies which relate to the measurement of science achievement. Our analyses should enable us to hypothesize about possible causes of any differences that are found, or alternatively to eliminate some potential causes from consideration.

Our paper begins with a rationale for the synthesis and brief discussions of measurement issues and of gender differences in science achievement. The methods for this review are next discussed, and are compared briefly to those used by Steinkamp and Maehr. Our findings and a discussion conclude the paper.

Rationale for the Review

Steinkamp and Maehr (1983) reported a quantitative synthesis of correlations among affect, ability, and achievement in science; and between each of these variables and gender. Their primary question in examining the literature on cognitive and attitudinal origins of science achievement was to determine whether science instruction should focus especially on affective outcomes.

Steinkamp and Maehr summarized correlation coefficients between gender and science achievement from 15 studies, using _t_ tests to examine the significance of the average correlations. One of Steinkamp and Maehr's conclusions was that boys achieve slightly better than girls in science, and they elaborated on that finding by calculating average correlations for samples grouped by the subject matter of the science test used and by the school grade level of the subjects in the sample. Steinkamp and Maehr did not, however, statistically examine the variation in their results.

Steinkamp and Maehrs' second review (1984) was primarily concerned with gender differences in motivational orientations toward science achievement, though they summarized results on gender differences in achievement as well. Using Glass's (1976) effect size to represent the extent of gender differences, the authors found that across 406 samples, boys' achievement averaged slightly better than that of girls. Though they reported averages and standard deviations for sex differences based on studies from different sources (i.e., articles versus standardized-test manuals), Steinkamp and Maehr did not focus on achievement gender differences in this review.

The variability in the results from both of Steinkamp and Maehr's reviews raises questions about their conclusions. One might ask whether the average correlations or effect sizes they report are representative of all their study results. That is, one could investigate whether the studies share one population correlation or one population effect size. If not, average correlations (or mean differences) could be misleading in that they would not accurately describe results in all (or perhaps any) of the studies. One could also examine the similarities or differences between the correlational results and the results from noncorrelational studies. More important, however, is the question of variation in the size of the gender differences. If we can find other variables which relate to the sizes of the sex differences, we may begin to understand reasons for and causes of any differences that do exist. These issues will guide our analyses in this paper.

Measurement and Gender Differences

We focus in this review on issues in the measurement of science achievement. Because the construct of science achievement is most often represented by the scores obtained by students on science achievement tests or the grades that they obtain in science classes, the instruments have much to do with the outcomes of quantitative reviews.

Several obvious features of the measures of science achievement seemed likely to relate to gender differences in science achievement. Though little work has been done in the area of science achievement, research in other content domains has shown that test length, test content (specifically within-subject differences in content such as geometry versus algebra in mathematics) , and test format may relate to gender differences in performance (Dwyer, 1979).

Test length and the speededness of a test have not been shown to consistently relate to gender differences, though Dwyer (1979) noted that speeded tests in masculine contexts have been found to favor males.

Some evidence (Finn, Dulberg, & Reis, 1979) suggests that gender differences cross-nationally relate to both test content and student age. Evidence from the study of the International Association for the Evaluation of Educational Achievement (IEA) suggested that gender differences were smaller for younger students and increased to almost a full standard deviation advantage for males by the end of secondary school. Boys were noted to excell the most in physical sciences, and less in biology. In some countries girls outperformed boys in biology.

Also, flaws or biases in the construction of the measures may relate to the size of the gender differences on the measures, though such relationships may be complex or difficult to discover without detailed data about the author(s) of the instrument and details about the method of construction.

In this analysis we examine test content, test length, and student grade, as well as some other features of the measurement and analysis of science achievement.

Meta-analysis of Gender Differences

Methods of integrative reviewing have developed much since Glass (1976) introduced the effect size and the idea of meta-analysis. Glass suggested the effect size, or standardized difference between the means of a treatment and a control group, as a scale-free measure of treatment effect. (In this review the effect size serves as a measure of gender differences.) Glass felt that this quantitative index could be used to combine study results in a quantitative and more objective fashion, following the procedures and rigor required of primary data analysis.

Hedges (1981) pointed out that Glass's meta-analysis is designed to draw inferences about a population effect size, which he called δ, through analyses of sample estimates. Early meta-analyses (including those by Steinkamp and Maehr) used intuitively sensible and familiar statistical analyses such as t tests and analysis of variance to analyze sample effect sizes. However, those methods are usually not appropriate for the analysis of effect-size data because they require the assumption of homogeneous variances across units (in this case, studies). Since effect sizes are based on mean differences, their variances, like the standard error of the sample mean, will depend on the sizes of the samples for which they are calculated. Thus, variances of effect sizes are usually heterogeneous.

Hedges (1982a) further indicated that when the effect sizes themselves are not similar a pooled or average estimate can be misleading. The original idea of Glass's meta-analysis was improved by Hedges's (1981) methods for distinguishing among effect sizes for studies which do not share a common (population) effect size. In our analyses we estimate several simple categorical (ANOVA-like) and regression-like models and examine whether any of them adequately explains variability in the sizes of the gender differences.

Methods

The Collection of Studies

Published studies examining sex differences in science achievement from both of Steinkamp and Maehr's earlier reviews (1983, 1984) are included in this review. A total of 120 distinct sources were identified from the bibliographies provided by Steinkamp and Maehr (1983, 1985). Though Steinkamp and Maehr also obtained effect-size estimates from test manuals and large scale studies they did not provide references for those sources. Thus the review was confined to the published sources of data listed in the bibliographies. Six dissertations and three unpublished manuscripts (not available through ERIC) were excluded from the review.

One hundred and eight published articles and ERIC documents were retrieved, and 42 of those sources were identified as having examined sex differences in science achievement. Each article was then read, effect-size measures were extracted from as many distinct independent samples as possible, and relevant study features were coded. (The original data from the Steinkamp and Maehr syntheses was requested, but was not available.)

One shortcoming of the data set examined here is that it does not contain many samples representing results from England. In particular, the work of Kelly (e.g., Kelly, 1978) and others on the Girls into Science and Technology project is an important piece of the research on science and gender. We plan to incorporate the data into our analyses as soon as it can be obtained.

Study Coding

Study features. Numerous study characteristics were coded for the 48 samples in the final collection. Table 1 presents a list of the study features used in our analyses.

Table 1

Features of Studies

Study feature	Categories
Subject-matter content	General science Biology Chemistry Geology/earth sciences Physics/physical science
Testing design	Analysis of covariance Posttest only Pretest/posttest (change)
Type of achievement measure	Constructed for the study Locally standardized Standardized Course grade
Number of items on measure	
School grade of subjects	

Of primary interest were the features of studies which related to measurement issues and to study design. Several features of the instruments used to measure science achievement were coded, including the subject-matter (content) tested, the reliability of the instrument, the test length, and the type of test which was used (e.g., standardized tests versus course grades).

The subject-matter content of the measures was determined through descriptions given in the studies and inspection of the instruments. Many articles described the tests in detail or presented all of the test items (or at least sample items). Even when detailed information was not available, titles often described the general content of the measures, and more specific information was sometimes available in test manuals or handbooks (for standardized tests especially). When a test contained a mixture of several specific science content areas (e.g., Bowyer & Linn, 1978) it was categorized as a test of general science. Several measures specifically labeled as tests of general-science ability (e.g., Field & Cropley, 1969) were also categorized into this grouping.

Additional features coded included the date of data collection and publication for each study, the nationality of the subjects, the source of the study (e.g., journal versus book), and several others.

Subjects of the studies were typically primary or secondary school students; only six samples of college students were included in the review. School grade level of the students was noted, and all college samples were assigned a grade of "13," since most often their actual college class level was not indicated.

<u>Effect sizes</u>. Sample sizes were obtained or estimated from all studies, and 30 effect sizes were extracted from samples which provided sufficient data for their calculation. We chose to represent the sex difference in the effect-size metric rather than as a correlation because effect sizes are more easily interpreted than correlation coefficients. That is, effect sizes directly represent the difference in means between the sexes. Furthermore, the distribution of the effect size is approximately

normal even when its population mean is nonzero, whereas the same is not true for the correlation.

Glass's (1976) effect size was calculated for each study. The mean difference between males and females on the science outcome measure (denoted here as Y) is used instead of the mean difference between treatment and control groups. A positive effect size represents a male advantage on the science achievement measure. The formula for Glass's effect size for the *i*th of a set of *k* studies is:

$$g_i = \frac{\bar{Y}^M{}_i - \bar{Y}^F{}_i}{S_i}, \tag{1}$$

where S_i is the pooled standard deviation from the usual two-sample *t* test for male and female groups. When only *t* or *F* statistics were presented in the studies, g was calculated as $\sqrt{(n_M + n_F)/n_M n_F}\ t$, which is algebraically equivalent to (1).

In many cases studies did not present the raw means and standard deviations needed to compute g as it is shown in (1). In such cases algebraic transfomations were used (e.g., Glass, McGaw, & Smith, 1981, pp. 93-152) to obtain g from available data. When analysis of variance summary statistics were presented, sums of squares for between-subjects terms (other than gender) were repooled into the error sum of squares to give an error estimate comparable to S_i.

When point biserial correlations of subject gender with the science outcome were reported, (e.g., Doran & Sellers, 1978), approximate g values were obtained via

$$g' = \frac{\bar{Y}^M{}_i - \bar{Y}^F{}_i}{S'} = \frac{(n_M + n_F)}{\sqrt{n_M \cdot n_M}}\, r_{GY}, \tag{2}$$

where r_{GY} is the point-biserial correlation between gender and the science outcome and n_M and n_F are the sample sizes for males and females. This effect size may differ from g in that its denominator S′ tends to be larger than S_p when gender differences are large. In these studies, however, it appears that the difference between S′ and S_p is negligible. The mean ratio of S′ to S_p in the ten samples in which both standard deviations could be computed was 1.012, which has little influence on the value of g′. Thus g′ values were considered equivalent to the g values computed as in (1).

Hedges (1981) found that Glass's estimator g has a small sample bias, and he obtained a corrected effect size d_i, which is the minimum variance unbiased estimator of δ. The unbiased estimator, corrected for small-sample bias and unreliability, is

$$d_i = (1 - 3/(4m_i - 1))\ g_i/\sqrt{r_i}$$

where m_i is $n_M + n_F - 2$ and r_i is the reliability estimate for the science achievement measure (Y) used in the ith study. When test reliability was not reported we used as r_i the average reliability for the rest of the studies, .82. We use the corrected effect sizes (d's) in our analyses.

Model Fitting and Estimation

The analyses reported in our paper are based on Hedges's (1982a,b) tests for fitting categorical models to effect sizes. Our procedure was to first ask whether the size of the gender difference on science achievement (across all the studies) was consistent. If the results were inconsistent, then studies were categorized according to one (or more) of the study features listed in Table 1. The agreement among results within each subset of results was then examined, as were possible differences between the groups.

When effect sizes within the groups appeared fairly similar, the analyses were stopped; if not, further subdivision of the groups continued until a sensible model was found or until the selected predictor variables were exhausted.

Our analyses differ from those of Steinkamp and Maehr because we not only examine between-study differences in the magnitudes of gender effects, but we follow with an examination of whether nonrandom variation remains within the sets of effects being considered. This allows us to address the question of whether an "adequate" explanation or model for the results has been found.

Results

Effect Sizes

We first tested the homogeneity of the results of the gender differences. Note that we omit one of the two effect sizes from Marjoribanks (1976) in order to have only independent data in the analysis. Table 2 shows the analysis of the set of 29 effects.

The homogeneity test H_T value was 101.39, which as a chi-square variable with $k - 1 = 28$ degrees of freedom, is quite large ($p < .001$). All the effect sizes can not be represented with one population parameter. This does not seem surprising since the biased uncorrected effect sizes ranged from -0.36 to 0.43.

The average effect size for all studies is estimated to be 0.16 standard deviations, which differs from zero ($p < .05$). This value is lower than that reported by Steinkamp and Maehr (1983). Their correlational studies produced an average correlation of 0.16, which corresponds approximately to an average effect size of 0.32. Though this indicates that males are on average achieving higher science scores than females, the value is an average and not a common effect size value. Some studies show more of an advantage for males and others show less; some may also show female superiority.

We next grouped the effects according to the subject-matter content of the achievement test. Table 2 also shows the homogeneity statistics obtained for this first categorical analysis (Hedges, 1982b). An overall test of the within-groups homogeneity, H_W, is the sum of the homogeneity values for each subgroup. Its value, 50.34, is significant at the .001 level (df=24). Thus there is still considerable variation in the sizes of the gender differences within the subject-matter subgroups. However, Table 2 shows that the results within the five subject-matter categories are, for the most part, consistent. Only the gender differences based on tests labeled as general science are not homogeneous. Thus most of the variation in the overall H_W statistic results from the differences among the studies of general science.

Table 2

Subject-Matter Differences Between Effect Sizes

Source	df	Test of Homogeneity	p value	Mean effect-size estimate (s.e.)
Total	28	101.39	<.001	0.16 (0.02) *
Between subject-matter groups	4	51.05	<.001	
Within groups	24	50.34	<.001	
General science	10	30.68	<.005	0.07 (0.05)
Biology	5	4.54	ns	0.14 (0.04) *
Chemistry	0	0.00	-	-0.12 (0.06)
Geology	4	6.79	ns	0.10 (0.06)
Physics	5	8.33	ns	0.35 (0.03) *

The test for differences between mean effect sizes for the subject-matter groups is given by H_B, which is also a chi-square variable, with 4 degrees of freedom (one less than the number of categories considered). We conclude that the five sets of gender differences have different population effect sizes, since $H_B = 51.05$ is significant.

Mean gender differences for two of the subject-matter groups were significantly greater than zero. The effects for studies of biology and physics both showed advantages for males, of 0.14 and 0.35 standard deviations, respectively. There are no significant differences between males and females on either geology or chemistry, though the single study of chemistry shows a female advantage which is significant with a lenient significance level of .10.

We should note here that the effect size value for physics achievement from the omitted Marjoribanks (1976) study (with a value of 0.00) does not conform with the results presented in Table 2. In fact, when the Marjoribanks sample is included in the analysis of physics outcomes, the within-group homogeneity test for physics increases to 28.00, which is highly significant as chi-sqare variable with 6 degrees of freedom. Also the estimated average effect size is reduced to 0.29 when this study is added. Nonetheless the effect size is still larger than those form the other subject-matter areas.

We next subdivide the general-science studies according to the school grade of the subjects. Subgroups were elementary schoolers (grades 1 through 6), junior-high students (grades 7 through 9), senior-high students (grades 10 through 12), and college students. (No linear relationship of grade to the gender difference was found, thus a categorization of grade was used to mesh with the subject-matter categorical analysis.)

The homogeneity statistics for studies of general science divided by grade are shown in Table 3. Only studies of junior-high groups share a common population effect size, which was significantly different from zero and indicates more than a quarter of a standard deviation advantage for males. The effect sizes on general science for elementary or senior-high subjects are still inconsistent, and are on average smaller than those for the junior-high subjects.

Table 3

Analysis of Gender Differences in General Science by Grade of Subjects

Source	df	Test of Homogeneity	p value	Mean effect-size estimate (s.e.)
General science	10	30.68	<.005	0.07 (0.05)
Elementary	5	13.70	<.02	0.02 (0.05)
Junior high	1	0.67	ns	0.29 (0.11) *
Senior high	2	9.54	<.01	0.17 (0.11)
Between Grade	2	6.77	<.05	

Since grouping the studies by grade for general science does not fully explain the variations in gender differences, we explored the use of another study feature as a grouping variable: the type of measure used. (The studies of general science had used standardized and locally-made tests, and tests made specifically for the study).

We find that the results of studies using standardized or locally made measures are consistent. However, studies using measures constructed specifically for the research have quite inconsistent results. Homogeneity statistics for the general-science studies grouped by the type of measure used are shown in Table 4.

Table 4

Analysis of Gender Differences in General Science by Type of Measure

Source	df	Test of Homogeneity	p value	Mean effect-size estimate (s.e.)
General science	10	30.68	<.005	0.07 (0.05)
Standardized	2	4.21	ns	-0.07 (0.12)
Local	0	0.00	ns	0.35 (0.14) *
Made for study	6	20.73	<.005	0.08 (0.06)
Between Measures	2	5.74	<.05	

Again we find that though results for some of the subgroups in the analysis are homogeneous, they still seem to vary considerably within one group. Here, results are still quite varied within the category of general science tests which were constructed specifically for the study in question.

Mean effects differed significantly between the three measure-type groups, with a small (though non-significant) superiority for girls found in studies using standardized tests. The one study using a locally standardized test showed over a third of a standard deviation advantage for males, though the generalizability of this finding is questionable since it is based on only one study.

Again, considerable variability in the results remains for one category of studies, studies using measures constructed specifically for the research project reported. The tests in this category included some with content from several domains within science, and varied in format as well. For example, the Scientific Literacy Test used by Bowyer and Linn (1978) was a pencil-and-paper test of content and process topics based on the goals of the Science Curriculum Improvement Study. On the other hand, the TAB Inventory of Science Processes used by Thomas and Snider (1969) was based on a sample of student behavior as he or she solved a science problem. Because the category "general science" was used, in a sense, as a catch-all category in the coding of the subject-matter content of the measures, the remaining amount of variation in the results is somewhat expectable.

We did no further subdivision of the studies because the number of remaining studies is quite small after the studies of general science are classed by grade level or by type of measures.

In order to examine the question of competing hypotheses as explanations for the gender differences in science achievement we considered several other categorical and regression-like models for the gender differences. Two other simple models involved classifying the studies according to either the grade level of the students or the type of measure used in the study. There were significant between-grade and between-measure differences, but there was also considerable variation within subgroups. For the measure-type analysis, though, all of the excess variation was concentrated in the largest subgroup of studies, those using author-made measures.

The predictors of publication date, study date, and test length did not appear related to the size of the gender difference, either in a linear or nonlinear fashion. Similarly, classifying the studies according to either the nationality of the subjects or the design of the study (e.g., posttest versus change-score analysis) did not produce a well-fitting explanatory model. Since the subject-matter predictor is the most salient predictor theoretically, it is reasonable that it provides the best-fitting model for the results.

Discussion

Gender differences for all subject-matter groupings except for studies of general science are consistent, and the average gender differences are all less than one half of a standard deviation. In physics and biology boys tend to do significantly better than girls, and the sizes of the gender differences are about one-third of a standard deviation for physics and about one sixth of a standard deviation for biology.

These results resemble those found by Finn, Dulberg, and Reis (1979), whose cross-cultural examination of the science achievement of adolescents showed smaller sex differences for biology than for other science areas. However, our gender differences do not appear to incresae with age and are much smaller in absolute magnitude than those reported by Finn, Dulberg and Reis. Some have suggested that the larger gender differences in physics result from girls' lack of mathematical skills. Unfortunately, our analyses are unable to shed any light on that matter.

Much of our analysis involved the examination of the studies of general science, and consideration of competing models to explain the gender differences across all studies. It is understandable that the results for studies of general science are inconsistent. Most of the general science measures seemed to vary internally in content and format. In part we had coded these measures as "general science" because they did not have specific subject-matter content. Differences due to variation in quality and content of the author-made measures also may have contributed to the heterogeneity of the general-science group.

It was interesting to find that other potential explanatory factors did not account for the variation in effect sizes as well as the subject-matter factor. Significant differences were found between some other groupings of studies (e.g., grade-level and measure-type groupings), though these categorizations left much variation unexplained.

Similarly, nationality of the students, study design, number of test items, and date of publication did not relate to the gender differences. Thus in our data the nationality of the students does not appear to be an important factor in explaining gender differences, and for the samples investigated here the length of the test did not relate to the size of the gender difference. In most cases, however, it was not possible to determine whether the test was speeded, so the question of whether gender differences are due to test speed (independent of test length) can not be addressed.

All of these findings suggest a few conclusions. One is that the degree of gender differences in achievement varies significantly across subject-matter areas in science. This suggests that care should be taken to distinguish between content areas when discussing or researching science achievement and gender.

Secondly, much of the variation in results appeared for tests constructed by study authors for the express purpose of their research. There are many possible sources of bias inherent in tests that have not been rigorously scrutinized as (presumably) is true of most standardized tests. Some author-made tests in this review do not appear to have been pilot-tested before their use in the published research. Though our present data did not permit such an analysis, a closer examination of the measures of science achievement, especially of general science achievement may be warranted.

References

Bowyer, J. B., & Linn, M. C. (1978). Effectiveness of the Science Curriculum Improvement Study in teaching scientific literacy. Journal of Research on Science Teaching, 15, 209-219.

Brickley, L. T., Garfunkel, G., & Hulsizer, D. (Eds.). (1979). Women and education [Special issue]. Harvard Educational Review, 49(4).

DeBoer, G. E. (1984). Factors related to the decision of men and women to combine taking science courses in college. Journal of Research on Science Teaching, 21, 325-329.

Doran, R. L., & Sellers, B. (1978). Relationships between students' self-concept in science and their science achievement, mental ability, and gender. Journal of Research on Science Teaching, 15, 527-533.

Dwyer, C. A. (1979). The role of tests and their construction in producing apparent sex-related differences. In M. A. Wittig & A. C. Petersen (Eds.). Sex-related differencesd in cognitive functioning. New York: Academic Press.

Finn, J. D., Dulberg, L., & Reis, J. (1979). Sex differences in educational attainment: A cross-national perspective. Harvard Educational Review, 49, 477-503.

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

Glass, G. V, McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

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

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

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

Hedges, L. V., & Olkin, I. (1981). Analysis, reanalysis, and meta-analysis. Comtemporary Education Review, 1, 157-165.

Kelly, A. (1978). Girls and science: An international study of sex differences in school achievement. Stockholm: Almqvist & Wiksell.

Marjoribanks, K. (1976). School attitudes, cognitive ability, and academic achievement. Journal of Educational Psychology, 68, 653-660.

National Science Foundation. (1977). Women and minorities in science and engineering. NSF Report No. 77-304. Washington DC: U.S. Government Printing Office.

Steinkamp, M. W., & Maehr, M. L. (1983). Affect, ability, and science achievement: A quantitative synthesis of correlational research. Review of Educational Research, 53, 369-396.

Steinkamp, M. W., & Maehr, M. L. (1984). Gender differences in motivational orientations toward achievement in school science: A quantitative synthesis. American Educational Research Journal, 21, 39-59.

Steinkamp, M. W., & Maehr, M. L. (1984). Personal communication.

Thomas, B., & Snider, B. (1969). The effects of instructional method upon the acquisition of inquiry skills. Journal of Research on Science Teaching, 6, 377-386.

Table 1

Results of the Studies of Science Achievement[1]

Study	Subject matter	Type of measure	Reliability	n_M	n_F	School grade	g	Number of items
Allen (1970)	General	Stdzd	0	150	150	1	6	999
Allen (1972)	General	Author	0	105	106	2	6	14
Allen (1973)	General	Author	0	100	76	3	0.42	25
Allen (1975)	General	Author	0	162	162	5	8	999
Anderson (1980)	Physics	Author	0	82	53	6	8	36
Ashbaugh (1968)	Geology	Author	.84	52	54	4	0.43	40
Ashbaugh (1968)	Geology	Author	.84	46	47	5	0.00	40
Ashbaugh (1968)	Geology	Author	.84	47	47	6	0.26	40
Babikian (1971)	Physics	Author	.76	108	108	8	0.36	38
Bowyer & Linn (1978)	General	Author	.91	284	247	6	-0.15	28
Bridgham (1969)	Physics	Author	66	29	21	3	0.42	30
Brown, Michaels & Bledsoe (1965)	Biology	Author	.90	112	111	13	7	60
Carnes, Bledsoe & VanDeventer (1967)	General	Author	.84	110	111	7	9	50
Clarke (1972)	General	Stdzd	0	415	361	5	7	999
Doran & Ngoi (1979)	Physics	Author	.64	101	101	6	9	999
Doran & Sellers (1978)	Biology	Unsure	.90	160	160	4	0.22	999
Field & Cropley (1969)	General	Local	0	104	74	11	0.38	999
Finger, Dillon & Corbin (1965)	Physics	Grade	0	137	42	13	8	999
Fuller, May & Butts (1979)	Biology	Author	.97	66	67	3	0.37	40
Hart (1978)	Biology	Author	.67	150	150	12	0.12	35

Table 1 (continued)

Results of the Studies of Science Achievement[1]

Study	Subject matter	Type of measure	Relia-bility	n_M	n_F	School grade	g	Number of items
Keeves (1975)	General	Local	0	107	108	7	0.32	999
Kempa & Ward (1975)	Chemistry	Author	.69	127	13	9	9	20
Kruglak (1970)	Physics	Stdzd	0	650	230	13	8	999
Lynch, Benjamin, Chapman, Holmes, McCammon, Smith, & Symmons (1979)	Physics	Author	.71	969	666	8	0.22	16
Lynch & Patterson (1980)	Physics	Author	0	969	666	8	6	16
Marek (1981)	Biology	Author	.73	37	55	10	0.18	102
Marjoribanks (1976)	Physics	Author	.94	201	195	6	-0.12	999
Marjoribanks (1976)	Biology	Author	.93	201	195	6	0.00	999
Marjoribanks (1978)	Biology	Author	.93	219	210	6	6	999
McDuffie & Beehler (1978)	General	Local	.82	196	197		-8	40
Ogden & Brewster (1977)	General	Stdzd	.88	63	20	11	0.28	60
Ogden & Brewster (1977)	General	Stdzd	.88	41	50	11	-0.36	60
Raven & Adrian (1978)	General	Stdzd	.85	132	117	10	7	999
Scott & Siegel (1965)	General	Author	.78	50	50	4	-0.05	20
Scott & Siegel (1965)	General	Author	.78	50	50	5	-0.18	20
Scott & Siegel (1965)	General	Author	.78	50	50	6	0.12	20
Shell (1970)	General	Stdzd	0	16	52	13	7	999
Shrigley (1972)	Earth Science	Author	.88	64	56	6	0.24	65
Sieveking & Savitsky (1969)	Chemistry	Stdzd	.96	498	498	13	-0.12	999
Skinner (1967)	Geology	Author	0	458	430	5	0.02	46
Strope & Braswell (1966)	General	Author	0	104	103	13	8	999

Table 1 (continued)

Results of the Studies of Science Achievement[1]

Study	Subject matter	Type of measure	Relia-bility	n_M	n_F	School grade	g	Number of items
Tamir (1974)	Biology	Local	0	259	256	12	8	999
Tamir (1974)	Biology	Local	0	259	256	12	-8	999
Tamir (1976)	Biology	Author	.79	468	521	12	0.12	30
Tamir & Amir (1975)	Physics	Author	.74	65	51	1	0.36	999
Tamir & Amir (1975)	Physics	Author	.72	65	51	2	0.16	999
Thomas & Snider (1969)	General	Author	.82	45	41	8	0.13	60
Walberg (1969)	Physics	Local	.77	675	375	12	0.42	999
Wallach & Kogan (1966)	General	Stdzd	.91	70	81	5	-0.03	999
Weisberg (1970)	Biology	Author	.84	48	48	8	7	40

[1] The codes for the variables are as follows:
Reliability: 0 = not reported, mean reliability of 0.82 was substituted.
Effect size (g): 6 = dependent; same subjects appear in another study.
7 = data for g is given, but direction is not reported.
8 = direction for g is given, but data is not reported (sign of 8 indicates direction of the difference).
9 = neither data nor direction of the difference is reported.
otherwise, the value of g is the computed effect size.
Number of items: 999 = missing data.

List of Synthesized Studies

Allen L. R. (1970). An evaluation of certain cognitive aspects of the material objects unit of the science curriculum improvement study elementary science program. Journal of Research in Science Teaching, 7, 277-281.

Allen, L. R. (1972). An evaluation of children's performance on certain cognitive, affective, and motivational aspects of the interaction unit of the science curriculum improvement study elementary science program. Journal of Research in Science Teaching, 9, 167-173.

Allen, L. R. (1973). An evaluation of children's performance on certain cognitive, affective, and motivational aspects of the systems and subsystems unit of the science curriculum improvement study elementary science program. Journal of Research in Science Teaching 10, 125-134.

Allen, R. (1975). Using a single topic film with elementary school children. Journal of Research in Science Teaching, 12, [illegible] 295.

Anderson, C. & Butts, D. (1980). A comparison of individualized and group instruction in a sixth-grade electricity unit. Journal of Research in Science Teaching, 17, 139-145.

Ashbaugh, A. C. (1968). Selection of geological concepts for intermediate grades. Science Education, 52, 189-196.

Babikian, Y. (1971). An empirical investigation to determine the relative effectiveness of discovery, laboratory, and expository methods of teaching science concepts. Journal of Research in Science Teaching, 8, 109-201.

Bowyer, J. B. & Linn, M. C. (1978). Effectiveness of the science curriculum improvement study in teaching scientific literacy. Journal of Research in Science Teaching, 15, 209-219.

Bridgham, R. G. (1969). Classification, seriation and the learning of electrostatics. Journal of Research in Science Teaching, 6, 118-127.

Brown, D. R., Michaels, G. E., & Bledsoe, J. C. (1965). An experiment in the use of film slides in an introductory course in microbiology. Journal of Research in Science Teaching, 3, 333-344.

Carnes, P. E., Bledsoe, J. C., & VanDeventer, W. C. (1967). Programmed materials in seventh-grade open-ended laboratory experiences. Journal of Research in Science Teaching, 5, 385-396.

Clarke, C. O. (1972). A determination of commonalities of science interests held by intermediate grade children in inner-city, suburban, and rural schools. Science Education, 56, 125-136.

Doran, R. L., & Sellers, B. (1978). Relationships between students' self-concept in science and their science achievement, mental ability, and gender. Journal of Research in Science Teaching, 15, 527-533.

Doran, R. L., & Ngoi, M. K. (1979). Retentention and transfer of selected science concepts in elementary school students. Journal of Research in Science Teaching, 16, 211-216.

Field, T. W., & Cropley, A. J. (1969). Cognitive style and science achievement. Journal of Research in Science Teaching, 6, 2-10.

Finger, J. A., Jr., Dillon, J. A., Jr., & Corbin, F. (1965). Performance in introductory college physics and previous instruction in physics. Journal of Research in Science Teaching, 3, 61-65.

Fuller, E. W., May, D. H., & Butts, D. P. (1979). The science achievement of third graders using visual, symbolic, and manipulative instructional treatments. Journal of Research in Science Teaching, 16, 129-136.

Hart, E. P. (1978). Examination of BSCS biology and nonbiology students' ecology comprehension, environmental information level, and environmental attitude. Journal of Research in Science Teaching, 15, 73-78.

Keeves, J. P. (1975). The home, the school, and achievement in mathematics and science. Science Education, 59, 439-460.

Kempa, R. F., & Ward, J. E. (1975). The effect of different modes of task orientation on observational attainment in practical chemistry. Journal of Research in Science Teaching, 12, 69-76.

Kruglak, H. (1970). Pre- and Post-sputnik physics background of college freshman-II. Journal of Research in Science Teaching, 7, 41-42.

Lynch, P. P., Benjamin, P., Chapman, T. , Holmes, R., McCammon, R., Smith, A., & Symmons, R. (1979). Scientific language and the high school pupil. Journal of Research in Science Teaching, 16, 351-357.

Lynch, P. P. & Paterson, R. E. (1980). An examination of gender differences in respect to pupils' recognition of science concept definitions. Journal of Research in Science Teaching, 17, 307-314.

Marek, E. A. (1981). Correlations among cognitive development, intelligence quotient, and achievement of high school biology students. Journal of Research in Science Teaching, 18, 9-14.

Marjoribanks, K. (1976). School attitudes, cognitive ability, and academic achievement. Journal of Educational Psychology, 68, 653-660.

Marjoribanks, K. (1978). The relation between students' convergent and divergent abilities, their academic performance, and school-related affective characteristics. Journal of Research in Science Teaching, 15, 197-207.

McDuffie, T. E., Jr., & Beehler, C. (1978). Achievement-workstyle relationships in ISCS Level I. Journal of Research in Science Teaching, 15, 485-490.

Ogden, W. R., & Brewster, P.M. (1977). An analysis of cognitive style profiles and related science achievement among secondary school students. ERIC document ED 139 610.

Raven, D. J., & Adrian, S. M. (1978). Relationships among science achievement, self-concept, and Piaget's operative comprehension. Science Education, 62, 471-479.

Scott, N. C., Jr. & Siegel, I. W. (1965). Effects of inquiry training in physical science on creativity. ERIC Document ED 003 700.

Shell, W. B. (1970). Interim testing with programmed material in college general science. Journal of Research in Science Teaching, 7, 245-248.

Sieveking, N. A., & Savitsky, J. S. (1969). Evaluation of an achievement test prediction of grades, and composition of discussion groups in college chemistry. Journal of Research in Science Teaching, 6, 374-376.

Skinner, R., Jr. (1967). Inquiry sessions: An assist for teaching science via instructional television in the elementary schools. Journal of Research in Science Teaching, 5, 346-350

Shrigley, R. L. (1972). Sex difference and its implications on attitude and achievement in elementary school science. School Science and Mathematics, 72, 789,793.

Strope, M. B. & Braswell, A. L. (1966). A comparison of factual teaching and conceptual teaching in introductory college astronomy. Journal of Research in Science Teaching, 4, 95-97.

Tamir, P. (1974). A comparative study of students' achievement in botany and zoology. Journal of Biological Education, 8, 333-342.

Tamir, P. (1976). Factors which influence student achievement in high school biology. Journal of Research in Science Teaching, 13, 539-545.

Tamir, P. & Amir, R. (1975). Teaching science to first and second grade pupils in Israel by the audio-tutorial method. Science Education, 59, 39-49.

Thomas, B. & Snider, B. (1969). The effects of instructional method upon the acquisition of inquiry skills. Journal of Research in Science Teaching, 6, 377-386.

Walberg, H. J. (1969). Physics, femininity, and creativity. Developmental Psychology, 1, 47-54.

Wallach, M. A. & Kogan, N. (1966). Modes of thinking in young children. New York: Holt, Rinehart and Winston, Inc.

Weisberg, J. S. (1970). The use of visual advance organizers for learning earth science concepts. Journal of Research in Science Teaching, 7, 161-165.