DOCUMENT RESUME

ED 236 232 TM 830 754

AUTHOR	Ziomek, Robert L.; Wilson, Mark
TITLE	A Proposed Additional Index to Glass' Effect Size Estimator with Application to Mastery Learning Experiments. Technical Report Series No. 3.
INSTITUTION	Des Moines Public Schools, IA. Dept. of Evaluation and Research.
PUB DATE	83
NOTE	47p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; *Effect Size; *Estimation (Mathematics); *Mastery Learning; *Measurement Techniques; *Outcomes of Education; Research Methodology
IDENTIFIERS	*Effect Size Estimator (Glass G V); Variability; *Variance (Statistical)

ABSTRACT

The emphasis of mastery learning instruction is unique in that concern is not only with raising the level of achievement but also with reducing the variability of achievement among the students instructed. The Glass effect size estimator, presently used an an index to measure the effectiveness of mastery learning experiments, is examined relative to these two conditions. This paper advances the thesis that the Glass effect size estimator represents only one dimension in the assessment of mastery learning experimental outcomes. The second dimension of effect, relevant to mastery learning outcomes, is the variability of the effect. The index to be proposed is directed at assessing this condition, in addition to the information provided by the effect size estimator. (Author)

Research Evaluation Development

A proposed additional index to Glass' effect size estimator with application to Mastery Learning Experiments.

Robert L. Ziomek
Mark Wilson

NO.3

Technical Report Series

Des Moines Public Schools
"Striving for excellence"

Department of Evaluation and Research

A PROPOSED ADDITIONAL INDEX TO GLASS' EFFECT SIZE ESTIMATOR WITH APPLICATION TO MASTERY LEARNING EXPERIMENTS

Robert L. Ziomek
Des Moines Independent Community School District
and
The University of Chicago

Mark Wilson
The University of Chicago
MESA Psychometric Laboratory

Department of Evaluation and Research
Des Moines Independent Community School District
1800 Grand Avenue
Des Moines, IA 50307

 1983

AUTHOR'S NOTE

This paper was conceived and written while the principal author was a visiting scholar in the Department of Education, MESA Psychometric Laboratory, at the University of Chicago. The author would like to express his gratitude to Professor Benjamin Bloom, Larry Hedges and Art Burke for their assistance and encouragement to pursue this idea.

ABSTRACT

The emphasis of mastery learning instruction is unique in that concern is not only with raising the level of achievement but also with reducing the variability of achievement among the students instructed. The Glass effect size estimator, presently used as an index to measure the effectiveness of mastery learning experiments, is examined relative to these two conditions. This paper advances the thesis that the Glass effect size estimator represents only one dimension in the assessment of mastery learning experimental outcomes. The second dimension of effect, relevant to mastery learning outcomes, is the variability of the effect. The index to be proposed is directed at assessing this condition, in addition to the information provided by the effect size estimator.

A PROPOSED ADDITIONAL INDEX TO GLASS' EFFECT SIZE ESTIMATOR WITH APPLICATION TO MASTERY LEARNING EXPERIEMENTS

Stimulated by the publication of Jacob Cohen's book, Statistical Power Analysis for the Behavioral Sciences (1969), and popularized by Glass and others (Glass, 1976; Smith and Glass, 1977; Rosenthal and Rubin, 1978; Kulik, Kulik and Cohen, 1979; Glass and Smith, 1979), the concept of effect size (ES) has received much attention as a practical means of assessing treatment effectiveness and as a tool for the quantitative synthesis of research. Of particular interest in this paper is the utilization of the Glass effect size estimator as an index to measure the effectiveness of individual mastery learning (ML) experiments. The emphasis of mastery instruction is unique in that the ML philosophy is concerned not only with raising the level of achievement but also with reducing the variability of achievement. "As we have proposed elsewhere (Block, 1974), if mastery approaches do in fact help more students to learn better than has traditionally been the case, the mastery-taught students should exhibit greater learning, as well as less variability in their learning, than nonmastery-taught students. That is, the mastery-taught students should not only learn better, they should learn more like one another" (Block and Burns, 1977).

Traditional instructional models are not specifically concerned, a priori, with the dual problems of increasing the level of achievement and concomitantly reducing the variability of performance as the major outcomes of instruction. In these models, concern is usually concentrated upon the overall level of treatment effectiveness (effect size) as measured by the differences in average level of achievement.

Benjamin Bloom believes that the outcome of reduced variance resulting from ML experiments is one useful index of what his theory of school learning promises (Bloom, 1976). Block and Burns (1977) in their review of 39 ML experiments involving 97 comparisons of average achievement test scores in a variety of subject areas report that mastery-taught students scored higher than nonmastery-taught students 89 percent of the time, and significantly higher 61 percent of the time. In addition, 26 of the 39 reviewed ML experiments reported variance data involving 80 comparisons. Of the 80 comparisons, mastery students exhibited less variability 74 percent of the time on achievement test scores than nonmastery-taught students. Upon computing the ES estimator for each of these comparisons, Block and Burns report that mastery-taught students, scored on the average, two-thirds of a standard deviation higher than nonmastery-taught students. Similar results are reported by Bloom (1976).

It is clear then that the dual importance accorded to level and variability of achievement within the ML context necessitates the development of indices reflecting these conditions. The measurement of levels of achievement, vis a vis effect size, is appropriate in "traditional" instructional experiments as noted previously. However, this measure represents only one dimension in the assessment of ML experimental outcomes. The second dimension of effect, relevant to ML experiments, is the "variability" of the effect. The index to be proposed is directed at assessing this condition in addition to the information provided by the ES estimator.

THE GLASS EFFECT SIZE ESTIMATOR

The ES estimator as proposed by Glass (1976) is defined as the difference between the means of the experimental ($\overline{X}_E$) and control groups ($\overline{X}_C$) divided by the standard deviation of the control group (S_C). The index is a pure, "dimensionless," number expressed in terms of the standard deviation metric. Simply stated, ". . . the effect size (ES) is some specific non-zero value in the population. The larger this value, the greater the degree to which the phenomenon under study is manifested" (Cohen, 1977). Thus, ES = 1.0, indicates that the experimental group mean exceeds the control group mean by one standard deviation; assuming normality, only 16 percent of the control group participants are higher than the average experimental participant. This interpretation is justified if the two groups compared are distributed normally on the outcome measure and have homogeneous variances. However, these assumptions are not tenable for an ML experiment where the conditions of non-homogeneous variances and non-normality are likely to result. (In a latter section of this paper, pertaining to the application of the proposed index to the results of actual ML experiments, a test of the non-homogeneity of variances is conducted to verify the outcome of this condition.)

Although Glass (1978) notes that the definition of ES appears uncomplicated, he acknowledges that heterogeneous variances pose substantial difficulties in its formulation and interpretation. He then proceeds to present his arguments for utilizing the control group standard deviation as the metric of standardization via the consideration of the following examples:

	Experimental	Control
Means	$\overline{X}_E = 52$	$\overline{X}_C = 50$
Standard Deviations	$S_E = 2$	$S_C = 10$
Basis of Standardization	Effect Size	
S_E	1.00	
S_C,	0.20	
$(S_E + S_C)/2$	0.33	

He states that the measure of ES could be calculated by using either S_E, S_C or a combination of both, but notes the huge differences resulting. The average of the two standard deviations should be dismissed in his words "as merely a mindless statistical reaction to a perplexing choice" (Glass, 1978), and argues that the two remaining estimates are correct.

In Glass' second example, he considers two experimental groups compared with a control with the following results:

	Method A	Method B	Control
Means	50	50	48
Standard Deviations	10	1	4

His argument follows:

> If effect sizes are calculated using the standard deviation of the "method," then ES_A equals 0.20 and ES_B equals 2.00 -- a misleading difference, considering the equality of the method means on the dependent variable. Standardization of mean differences by the control group standards deviation at least has the advantage of allotting equal effect sizes to equal means. This seems reason enough to resolve the choice in favor of the control group standard deviation (Glass, 1978).

He further notes, however, that the problem of heterogeneous variances remains and has no clear answer.

In the preceding example, if the mean of the control group happened to equal 50, it is clear that the ES estimator would equal zero in both instances, implying no differences between the groups. This is a perplexing problem for measuring the effectiveness of mastery learning outcomes. If Method B represented a ML group, intuition would suggest a difference favoring that condition, i.e., although the level of learning is the same for all three groups, those individuals in Method B are learning, as a group, in a more homogeneous manner than those in Method A and the Control. The problem here is one of accounting for the presence of heterogeneous variances, when Glass' estimate of effect size is zero.

A PROPOSED SOLUTION

The effect variability index has the property that it utilizes all the descriptive statistics resulting from an experiment, i.e., the means and standard deviations of the groups to be compared, and does not go to zero in the presence of equal means associated with heterogeneous or homogeneous variances. Its utility will be illustrated by comparing it against the results of the two previously cited examples, in addition to applying it to actual ML experimental results to illustrate its interpretive power and ease of application.

The proposed index is derived by creating for each experimental group the coefficient of variation, i.e., the ratio of the sample standard deviation to the sample mean (Snedecor and Cochran, 1980). The index is then formed by etablishing the ratios of the coefficients of variation for the groups to be compared. That is,

$$\begin{aligned} I_{i/j} &= CV_i/CV_j, \quad i = j \qquad\qquad i = j = 1 \ldots k \text{ treatment groups} \qquad (1) \\ &= (s_i/\overline{X}_i)/(s_j/\overline{X}_j) \\ &= (s_i/s_j)\cdot(\overline{X}_j/\overline{X}_i) \end{aligned}$$

The impetus for this formulation is that each coefficient of variation reflects the amount of variability associated with a given mean, and subsequently, the ratio of these coefficients facilitates the assessment of their combined effectiveness. The interpretation of this index can best be illustrated by referring to Glass' first example:

	Experimental	Control
Means	$\overline{X}_E = 52$	$\overline{X}_C = 50$
Standard Deviations	$S_E = 2$	$S_C = 10$
Coefficients of Variation	2/52 = .038	10/50 = .20
ES (using control group SD)	0.20	

Recalling the definition of $I_{i/j}$,

$$I_{E/C} = (CV_E/CV_C)$$

$$I_{E/C} = (S_E/S_C)\cdot(\overline{X}_C/\overline{X}_E)$$

and substituting,

$$I_{E/C} = (2/10)\cdot(50/52)$$

$$I_{E/C} = (.2)\cdot(.96)$$

$$I_{E/C} = .19$$

The importance of this index is not necessarily reflected in the final number, but in the information provided by considering the factors determining the product. Each factor presents the relative impact of the individual ratio of the means and standard deviations. Let us explore this result in the context of an ML experiment.

The ES estimator equals 0.20, suggesting that the two groups do not differ greatly. Although this estimate may result in a satistically significant difference in the traditional sense, the magnitude of the number is considered a "small" effect within the context of ML experiments (Block and Burns, 1977; Bloom, 1976). In spite of the levels of achievement being relatively similar, the argument could be advanced that the lower achieving ML (Experimental) students are for the most part all achieving at a high level compared to the lower achieving students in the control group ($S_E = 2$ vs $S_C = 10$).

Although the two indices appear similar, ES = .20 and $I_{E/C}$ = .19, their meanings are different. The closer $I_{E/C}$ is to zero the greater the impact on variability of the experimental treatment; (i.e., the group represented in the numerator); whereas, the larger the value of ES (either in a positive or negative direction) the greater the effect favoring either the experimental or control group. As $I_{E/C}$ approaches one, the difference between the two groups becomes minimal. As $I_{E/C}$ exceeds one, the effect variability tends to favor the control group (i.e., the group represented in the denominator). Finally, where the ES estimator disappears in the presence of equal experimental means and heterogeneous or homogeneous variances, the proposed estimator retains this information -- the ratio of the means is simply one and the index is soley dependent on the ratio of standard deviations.

Let us illustrate this last situation by assuming that the control group mean in the present example equals 52. The ES estimator equals zero (no difference in experimental effect) whereas $I_{E/C}$ = .20, essentially equivalent to $I_{E/C}$ = .19, in which $\overline{X}_E$ = 52 and $\overline{X}_C$ = 50; a result primarily due to the approximate equality of the means compared, i.e., $\overline{X}_C/\overline{X}_E$ = 50/52 = .96, whereas $\overline{X}_C/\overline{X}_E$ = 52/52 = 1.0. Thus, the effect of heterogeneous variances is clear in the presence of equal or nearly equal means.

It is noted that the formulation of this ratio does not necessitate that the coefficient of variation of the experimental group appear in the numerator. For example, we could use $I_{C/E}$, then,

$$I_{C/E} = (S_C/S_E)\cdot(\overline{X}_E/\overline{X}_C) \qquad (2)$$

and substituting yields,

$$I_{C/E} = (10/2)\cdot(52/50)$$

$$I_{C/E} = (5)\cdot(1.04)$$

$$I_{C/E} = 5.2 = 1/I_{E/C}$$

Since the index exceeds one, the results favor the experimental group, i.e., the group represented in the denominator. Thus, $I_{E/C}$ and $I_{C/E}$ are reciprocally related. This relationship together with a comparison of $I_{E/C}$ and ES are summarized in Table 1. As is true for all newly proposed quantitative indices, terminology such as "close to zero" or "close to one" carry a subjective interpretation, eventually refined by familiarity with having employed the index in

TABLE 1. Characteristics of $I_{E/C}$

its	Interpretation	Equivalent Condition	Conditions	Interpretations: $I_{E/C}$	Interpretations: ES
approaches zero	experimental group favored	$I_{C/E}$ approaches one	I. $S_E > S_C$ & $\bar{X}_C > \bar{X}_E$	$.75 \leq I_{E/C} \leq 1.0$, no difference	small & negative, no difference
approaches one	no difference between groups	$I_{C/E}$ approaches one		$I_{E/C} < .75$, treatment favored	large & negative control favored
			II. $S_E < S_C$ & $\bar{X}_C < \bar{X}_E$	$.75 \leq I_{E/C} \leq 1.0$,	small & positive no difference
greater than one	control group favored	$I_{C/E}$ approaches zero		$I_{E/C} < .75$, treatment favored	large & positive treatment favored
			III. $S_E > S_C$ & $\bar{X}_C > \bar{X}_E$	$I_{E/C}$ close to one, no difference;	small & negative no difference
				"significantly" larger than one, control favored	large & negative control favored
			IV. $S_C > S_C$ & $\bar{X}_C < \bar{X}_E$	$.75 \leq I_{E/C} \leq 1.0$, no difference;	small & positive no difference
				"significantly" larger than one control favored	large & positive treatment favored

a number of experimental situations in addition to studying the distribution of the index (see Appendix B, derivation of the large sample normal approximation for $I_{E/C}$). For purposes of this discussion we will adopt the convention that close to zero implies, $0 \leq I_{E/C} \leq .35$; close to one will mean, $.75 \leq I_{E/C} \leq 1.0$; and $I_{E/C}$ between .35 and .75 will be considered an "average" effect.

Recall Glass' second example:

	Method A	Method B	Control
Means	50	50	48
Standard Deviations	10	1	4

Applying the Glass estimator to these results, recalling that the standardization metric is $S_C = 4$, ES is equal to .5 for both methods, suggesting both conditions have equal effectiveness (recall that this is effectiveness relative to the level of achievement and not effectiveness relative to the reduction of variance), in spite of heterogeneous variances. If one were to use the ES method to estimate the relative effectiveness between Methods A and B, regardless of the standard deviation (i.e., $S_A = 10$ or $S_B = 1$), ES would equal zero suggesting no difference. A close inspection of the results of this experiment, once again from a ML viewpoint, suggests that Method B is more effective than Method A.

a number of experimental situations in addition to studying the distribution of the index (see Appendix B, derivation of the large sample normal approximation for $I_{E/C}$). For purposes of this discussion we will adopt the convention that close to zero implies, $0 \leq I_{E/C} \leq .35$; close to one will mean, $.75 \leq I_{E/C} \leq 1.0$; and $I_{E/C}$ between .35 and .75 will be considered an "average" effect.

Recall Glass' second example:

	Method A	Method B	Control
Means	50	50	48
Standard Deviations	10	1	4

Applying the Glass estimator to these results, recalling that the standardization metric is $S_C = 4$, ES is equal to .5 for both methods, suggesting both conditions have equal effectiveness (recall that this is effectiveness relative to the level of achievement and not effectiveness relative to the reduction of variance), in spite of heterogeneous variances. If one were to use the ES method to estimate the relative effectiveness between Methods A and B, regardless of the standard deviation (i.e., $S_A = 10$ or $S_B = 1$), ES would equal zero suggesting no difference. A close inspection of the results of this experiment, once again from a ML viewpoint, suggests that Method B is more effective than Method A.

PROPERTIES OF THE EFFECT VARIABILITY INDEX

It can be shown that $I_{E/C}$ can be expressed as a linear function of ES (see Appendix A for derivation); i.e.,

$$I_{E/C} = CV_E/CV_C = S_E/S_C - (CV_E)\cdot(ES) \qquad (3)$$

Thus when ES = 0 (equivalent to $\overline{X}_E = \overline{X}_C$), $I_{E/C} = S_E/S_C$, is simply a function of the ratio of standard deviations. Consequently, the meaning of $I_{E/C}$ is apparent in the presence of equal treatment means and potentially heterogeneous variances.

In the reviewers notes concerning an earlier version of this paper (Glass, Shephard and Smith, 1983), it was noted that in the presence of <u>equal</u> standard deviations, $S_E = S_C$, the $I_{E/C}$ index is not invariant under a linear transformation[1]. The following example will serve to illustrate this condition. Assume $S_E = S_C = 5$ and $\overline{X}_C = 45$ and $\overline{X}_E = 55$. Then $I_{E/C} = .82$ and ES = 10. If the means were transformed by adding 10 to each, then $I_{E/C} = .85$, whereas ES remains constant, i.e., ES = 10. If the sample sizes for both the control and treatment groups were, $n_E = n_C = 50$, and applying the large sample normal approximation results discussed in Appendix B, both the initial index and the transformed index yield a nonsignificant difference between the treatment and control. The larger the additive constant the greater the difference between the original $I_{E/C}$ and the transformed estimate of effect variability.

We must ask how meaningful this problem is in a ML context. Obviously, for experiments using the same measuring instrument (i.e. tests), there is no such transformation, and so there is no problem. If the instruments used are similar (in metrical properties) then the problem will be small. If, however, the instruments are considerably different, then it would be possible to attain differences in $I_{E/C}$ that looked important, but were due only to the particular choice of instrument. It is the author's contention that the problem here is in the use of inappropriate measures and not in the index itself. One way to circumvent such problems would be to ensure that all instruments used in a series of ML experiments could be located on a common scale (one could use a Rasch linking procedure, outlined in Best Test Design, Wright and Stone 1979); it is hard to imagine an ML experiement in which this would not be good advice in any case. If one is comparing across experimental studies, however, and the instruments used in each differ markedly in their level of difficulty, then this problem remains. Such problems as these are not uncommon in indices and yet have practical application. For example, Hedges (1981) has shown that ES itself is biased even under homogeneous variances; however, this can be controlled and even ignored, so long as one is aware of the limitation in application.

In the presence of both unequal means and standard deviations, the situation improves. Assume $S_E = 5$ and $S_C = 10$, $\overline{X}_E = 55$ and $\overline{X}_C = 45$. Then $I_{E/C} = .41$ whereas ES = 10. Transforming both means by adding 10 to each yields $I_{E/C} = .42$, a slight increase of .01. Finally in the presence of equal means and unequal standard deviations, $I_{E/C}$ remains invariant, as illustrated by Equation 3. In general if $X' = aX + b$ then

$$I'_{E/C} = (S_E/S_C)\cdot[(a\overline{X}_C + b)/(a\overline{X}_E + b)] \qquad (4)$$

and $I_{E/C} \simeq I'_{E/C}$ if and only if, $b/a < \overline{X}_C$ and $b/a < \overline{X}_E$. The scaling parameter, a, does not change the ratio of standard deviations.

Contingent upon the previous discussion, it appears appropriate to recommend that $I_{E/C}$, the effect variability index, be used as a tool for assessing experimental outcomes <u>within</u> a given research study as opposed to between studies, relative to meta analysis concerns. However, if tests used in a variety of experimental situations are equivalent in their level of difficlty, the $I_{E/C}$ estimator is a suitable tool to employ.

APPLICATIONS TO REAL DATA

The reader is reminded that the consideration of $I_{E/C}$ has been stimulated by experiments associated within the ML context, where level and variability of achievement are both of concern as measures of effective ML instruction, a problem not adequately accounted for by the Glass estimator. The data examined in Table 2 are from a study undertaken by Block (1972) in which the criterion level of mastery was selectively varied across four treatment groups (95, 85, 75 and 65 percent mastery levels) and the results compared to a control, non-mastery group. Block noted "that as the performance level attained increased, average achievement test scores rose and the dispersion fell. The 85 and 95 percent mastery treatments not only helped students to attain significantly higher average scores than the non-mastery treatment, but the treatments also helped homogenize student achievement around these scores."

For the 75 and 65 percent mastery learning groups, the ES estimator is approximately zero; whereas $I_{75/C}$ = .53 and $I_{65/C}$ = .69 and "average" effect according to the convention adopted earlier. For the 95 and 85 percent mastery groups, ES = .64 and ES = .46 respectively; whereas, $I_{95/C}$ = .31 and $I_{85/C}$ = .39, effects strongly favoring the two ML groups. These differences are due primarily to the impact of the ratios of the standard deviations favoring the

TABLE 2. Comparisons Between I_{EC} and ES for Block Data

Treatment Group	Average Test Scores (Percent Correct)	Standard Deviation	(S_M/S_{NM}) ×	$(\bar{X}_{NM}/\bar{X}_M)$ =	$I_{E/C}$	ES
Mastery						
95 percent (N = [illegible]1)	64.9	9.09*	.41	.77	.31	.64
85 percent (N = 14)	60.7	10.49*	.47	.83	.39	.46
75 percent (N = 14)	50.8	11.79*	.53	1.00	.53	.01
65 percent (N = 12)	49.0	15.49	.69	1.00	.69	-.07
Non-mastery (N = 25)	50.5	22.40				

*Significant difference ($p < .05$) between treatment group standard deviation compared to control group.

experimental groups. Of the four comparisons permitted by the ES method, the 95, 85 and 75 percent mastery treatment groups' standard deviations are significantly different from the non-mastery group standard deviations. (The F-test for the ratio of two variances was employed.) For the 75 percent mastery group comparison, ES = .01, suggesting no difference between the treatment and control group; whereas $I_{E/C}$ = .53, implying a difference between the groups compared in the presence of heterogeneous variances.

Table 3 presents the results of the matrix of all possible comparisons among the treatment groups along with the confidence intervals for each calculated $I_{E/C}$ (see Appendix B for confidence interval formula). For example, reading down the 95 percent column, in which the 95 percent mastery group is compared to all other groups, each confidence interval containing 1.0 indicates a non-significant difference between the groups compared. The ES estimator does not permit comparisons among the treatment groups, whereas the $I_{E/C}$ estimator, together with the confidence intervals, implies that no differences exist among the 95, 85, and 75 percent groups; between the 75 and 65 percent groups; and between the 65 percent and non-mastery group. All other differences between the treatment groups and the non-mastery group are significant. Once again although the ES = .01 suggests no difference for the 75 vs non-mastery group comparison, the $I_{E/C}$ estimator does indicate a difference between these groups.

TABLE 3. Matrix of Comparisons of $I_{E/C}$ for Block Data

		Treatment Groups 95	85	75	65	ES Estimator
	95					.64*
Treatment Groups	85	81 (.20, 1.42)[1]				.46*
	75	.60 (.14, 1.06)	.74 (.21, 1.27)			.01*
	65	.44 (.09, .79)	.55 (.13, .97)	.73 (.17, 1.29)		-.07
Nonmastery		.32 (.09, .55)	.39 (.13, .65)	.52 (.17, .87)	.71 (.20, 1.22)	

*Significant difference ($p < .05$) between treatment group standard deviation compared to control group.

1. See Note. All confidence intervals are computed at the 99 percent level.

The second set of comparisons, presented in Table 4, come from a number of mastery learning experiments reported by Bloom (1976). The data have been rank ordered relative to the ES estimator and separated into two groups, heterogeneous and homogeneous variances. All confidence intervals for $I_{E/C}$ not containing 1.0 imply a significant difference favoring the ML group. Of the 8 comparisons within the heterogeneous group, all $I_{E/C}$ estimates are significantly different from 1.0. Of the 11 comparisons for the homogeneous group only 2 $I_{E/C}$ estimates are significant, favoring the ML group. It appears that in the presence of heterogeneous variances, the $I_{E/C}$ estimator parallels the ES index suggesting an effect favoring the ML group; that is, not only is effect size large relative to level of achievement (ES estimator) but also large relative to the reduction of variability ($I_{E/C}$ estimator) (This condition is true for all cases with one exception, noted in the Block data in Table 2, for ES = .01 and $I_{75/NM}$ = .52.). However, the "magnitude" of the effect implied by the $I_{E/C}$ estimates and the corresponding confidence intervals are not as strong as those implied by the ES index based upon the previously suggested guidelines. In the presence of homogeneous variances, the results are reversed. For the 8 ES estimators greater than the absolute value of .49, only one $I_{E/C}$ index is significant, compared to the large effect sizes reported. Consequently, in the presence of homogeneous variances, although

TABLE 4: Comparisons Between $I_{E/C}$ and ES for Bloom Data

$\overline{X}_E$	S_E	N_E	$\overline{X}_C$	S_C	N_C	$I_{E/C}$	$I_{E/C}$ Confidence Interval	ES Estimator
Heterogeneous Variances								
74.00	18.40	1985	55.30	19.10	1271	.72	(.67, .77)*	.98
60.70	10.10	15	46.50	15.00	93	.52	(.25, .79)*	.95
69.30	17.30	1723	49.50	22.30	1310	.55	(.51, .59)*	.89
71.40	21.40	1895	52.80	21.10	1410	.75	(.69, .80)*	.88
78.20	4.50	113	72.40	6.70	113	.62	(.47, .77)*	.87
73.50	19.00	1806	54.30	22.30	1104	.63	(.58, .68)*	.86
3.52	0.50	98	2.51	1.40	75	.26	(.17, .34)*	.72
90.30	5.20	20	85.90	10.10	21	.49	(.20, .76)*	.44
Homogeneous Variances								
30.30	3.80	30	26.10	4.70	21	.70	(.33, 1.07)	.89
64.40	16.30	19	52.80	13.20	24	1.01	(.41, 1.61)	.88
74.60	12.40	33	64.40	12.60	20	.85	(.39, 1.29)	.81
68.50	13.90	26	56.90	16.00	33	.72	(.36, 1.08)	.72
120.30	23.90	17	99.00	35.40	17	.56	(.18, .94)*	.60
73.10	12.10	168	66.80	11.90	92	.93	(.70, 1.16)	.53
3.46	0.59	24	3.07	0.79	22	.67	(.29, 1.05)	.49
77.10	12.00	20	71.40	14.70	17	.76	(.29, 1.23)	.39
61.30	8.10	9	56.10	14.10	9	.52	(.06, .98)*	.37
13.50	4.50	34	15.90	5.60	37	.94	(.49, 1.39)	-.43
59.00	14.80	31	67.40	17.00	30	.99	(.50, 1.48)	-.49

*Significant difference between the ML and control group. All confidence intervals are calculated at the 99 percent level.

the ES estimator is large relative to the level of achievement differences, the reduction in the variability is minimal.

Table 5 presents the results of Table 4 in a different prespective, emphasizing the importance of examining the component ratios of means and standard deviations, in order to clarify the results of the preceding discussion. The format of Table 5 parallels that of Table 4. The first apparent trend is that the component ratios of the means and standard deviations for the homogeneous variance group of comparisons exceed those of the heterogeneous variance group, indicating that the means and standard deviations of the ML and control groups are approximately equal numerically in the homogeneous group. Of the two significant $I_{E/C}$ indices in the homogeneous group, one apparent explanation for their significance is implied by the two smallest standard deviation ratios, .68 and .57 associated with nearly equivalent mean ratios of .82 and .92 respectively. That is, in the case of S_E/S_C = .57, the ML groups' standard deviation is only 57 percent of the control group's standard deviation; and in the presence of a mean ratio $\overline{X}_C/\overline{X}_E$ = .92, results in significance. A similar occurrence is exhibited in the heterogeneous group of comparisons for S_E/S_C equal to .67 and .51 relative to $\overline{X}_C/\overline{X}_E$ equal to .93 and .95 respectively. Consequently, the interactions of the two component ratios comprising $I_{E/C}$ do contribute to the interpretation and understanding of $I_{E/C}$. Simulation studies are

TABLE 5. Comparisons of the Component Ratios of the $I_{E/C}$ Index

S_E/S_C	$\overline{X}_C/\overline{X}_E$	$I_{E/C}$	ES Estimator
Heterogeneous Variances			
.96	.75	.72	.98
.67	.77	.52	.95
.78	.71	.55	.89
1.01	.74	.75	.88
.67	.93	.62	.87
.85	.74	.63	.86
.36	.71	.26	.72
.5[illegible]	.95	.49	.44
Homogeneous Variances			
.81	.86	.70	.89
1.23	.82	1.01	.88
.98	.86	.85	.81
.87	.83	.72	.72
.68	.82	.56	.60
1.02	.91	.93	.53
.75	.89	.67	.49
.82	.93	.76	.39
.57	.92	.52	.37
.80	1.18	.94	-.43
.87	1.14	.99	-.49

directed at investigating the effects of sample size relative to the sampling variance of $I_{E/C}$ are reported in Appendix B.

SUMMARY

The index proposed in this paper is advanced as an additional measure of treatment effect, specifically applied to experimental programs where both level and variability of the variable under study is of dual concern. The proposed effect variability estimator, compared to the ES estimator and presently used to measure the effect of ML programs:

1. utilizes the means and standard deviations of all treatment conditions to be compared;
2. does not go to zero in the presence of equal means and heterogeneous or homogeneous variances;
3. provides for enhanced interpretation and understanding via a straightforward examination of the component ratios and their relative impact;
4. facilitates the rank ordering of treatment effects within experimental studies among all treatment groups;
5. emphasizes the dual nature of "effect size" relative to both level of achievement and reduction of variability within the context of ML experiments.

The importance associated with both level and variability of achievement, especially relevant to ML experiments, demands the development of techniques to capture and illuminte these conditions. This paper is an attempt to resolve a preplexing problem. At present work is underway, directed at investigating the distributional properties of $I_{E/C}$ and comparing the power of $I_{E/C}$ relative to ES under the conditions discussed.

The importance associated with both level and variability of achievement, especially relevant to ML experiments, demands the development of techniques to capture and illuminte these conditions. This paper is an attempt to resolve a preplexing problem. At present work is underway, directed at investigating the distributional properties of $I_{E/C}$ and comparing the power of $I_{E/C}$ relative to ES under the conditions discussed.

REFERENCES

Block, J. H. Student learning and the setting of mastery performance standards. Educational Horizons, 1972 (Summer), 183-191.

Block, J. H., & Burns, R. B. Mastery learning. In L. S. Shulman (Ed.), Review of research in education (Vol. 4). Itasca, IL: Peacock, 1977.

Bloom, B. S. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Cohen, J. Statistical power analysis for the behavioral sciences (Rev. ed.). New York: Academic Press, 1977.

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

Glass, G. V. Integrating findings: The meta-analysis of research. In L. S. Shulman (Ed.), Review of research in education (Vol. 5). Itasca, IL: Peacock, 1978.

Glass, G. V., & Smith, M. L. Meta-analysis of the relationship between class-size and achievement. Educational Evaluation and Policy Analysis, 1979, 1, 2-16.

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

International Mathematical and Statistical Libraries, Inc. IMSL Library 1, (7th ed.) Houston, TX: Author, 1977.

Kulik, J. A., Kulik, C. C., & Cohen, P. A. A meta-analysis of outcome studies of Keller's personalized system of instruction. American Psychologist, 1979, 34, 307-318.

Rao, C. R. Linear statistical inference and its applications (2nd ed.). New York: Wiley, 1973.

Rosenthal, R., & Rubin, D. B. Interpersonal expectancy effects: The first 345 studies. Behavioral and Brain Sciences, 1978, 3, 377-415.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. American Psychologist, 1977, 32, 752-760.

Snedecor, G. W., & Cochran, W. G. Statistical methods (7th ed.). Ames, IA: The Iowa State University Press, 1980.

Wright, B. and Stone, M. Best test design. University of Chicago, MESA Press, 1979.

APPENDIX A

Derivation of $I_{E/C}$ as a Linear Function of ES

$$ES = \frac{\bar{X}_E - \bar{X}_C}{S_C}$$

$$ES = \left[\frac{\bar{X}_E - \bar{X}_C}{S_C}\right] \cdot \left[\frac{S_E}{S_E}\right]$$

$$ES = \bar{X}_E \cdot \left[\frac{1 - \bar{X}_C/\bar{X}_E}{S_E}\right] \cdot \frac{S_E}{S_C}$$

$$ES = \frac{\bar{X}_E}{S_E} \cdot \left[1 - \frac{\bar{X}_C}{\bar{X}_E}\right] \cdot \frac{S_E}{S_C}$$

$$ES = \frac{\bar{X}_E}{S_E} \cdot \left[\frac{S_E}{S_C} - \frac{\bar{X}_C}{S_C} \cdot \frac{S_E}{\bar{X}_E}\right]$$

$$\frac{S_E}{\bar{X}_E} \cdot (ES) = \frac{S_E}{S_C} - (CV_E)/(CV_C)$$

$$I_{E/C} = CV_E/CV_C = \frac{S_E}{S_C} - (CV_E) \cdot (ES)$$

APPENDIX B

Derivation of the Large Sample Normal Approximation for $I_{E/C}$

LARGE SAMPLE NORMAL APPROXIMATION

The derivation of the large sample normal approximation for $I_{E/C}$ appeals to the delta (δ) method discussed in Rao (1973): Let T_n be a k-dimensional statistic $(T_{1n},\ldots T_{kn})$ such that the asymptotic distribution of $\sqrt{n}(T_{in} - \theta_i)$, i = 1...k, is k-variate normal with mean zero and dispersion matrix $\Sigma = (\sigma_{ij})$. Let $g_1,\ldots g_q$ be q functions of k-variates and each g_i be totally differentiable. Then the a.d. of $\sqrt{n}[g_i(T_{1n},\ldots T_{kn})-g_i(\theta_1,\ldots\theta_k)]$, i = 1, ...,q, is q-variate normal with zero means and dispersion matrix $G\Sigma G'$, where $G = (\delta g_i/\delta\theta_j)$. For $I_{E/C}$, let

$$\underset{\sim}{T} = \begin{pmatrix} S_1^2 \\ S_2^2 \\ \bar{X}_2 \\ \bar{X}_1 \end{pmatrix} \quad \text{and} \quad \underset{\sim}{\theta} = \begin{pmatrix} T_1^2 \\ T_2^2 \\ u_2 \\ u_1 \end{pmatrix} \qquad (1)$$

where $\underset{\sim}{T} \sim N(\underset{\sim}{\theta}_1, \Sigma)$, and assume that the asympototic distribution of $\sqrt{n}(\underset{\sim}{T} - \underset{\sim}{\theta})$ is $N(0, \Sigma)$ where Σ is defined as

$$\Sigma = \begin{pmatrix} 2T_1^4 & 0 & 0 & 0 \\ 0 & 2T_2^4 & 0 & 0 \\ 0 & 0 & T_2^2 & 0 \\ 0 & 0 & 0 & T_1^2 \end{pmatrix} \qquad (2)$$

Define the function g to be $g = \sqrt{T_1/T_2} \cdot (T_3/T_4)$, then the δ-method method yields

$$\sqrt{n}\,[g(\underset{\sim}{T})-g(\underset{\sim}{\theta})] \sim N(\underset{\sim}{0}, A\Sigma A') \qquad (3)$$

where $A = (a_{ij})$ and $a_{ij} = \delta g/\delta T_i$ evaluated at $\underset{\sim}{T} = \underset{\sim}{\theta}$. Taking the partial derivatives of $g(\underset{\sim}{T})$ and calculating $A\Sigma A'$ yields

$$A\Sigma A' = g^2(\underset{\sim}{T}) \left\{ \frac{1}{2}\,\theta_1{}^2 T_1{}^{-2} + \frac{1}{2}\theta_2{}^2 T_2{}^{-2} + \theta_2 T_3{}^{-2} + \theta_1 T_4{}^{-2} \right\} \qquad (4)$$

Evaluating $A\Sigma A'$ at $\underset{\sim}{T} = \underset{\sim}{\theta}$ and recalling that ($\theta_1 = T_1{}^2$, $\theta_2 = T_2{}^2$, $\theta_3 = u_2$ and $\theta_4 = u_1$) yields

$$\sigma_{\cdot}^2 = A\Sigma A' = (\sigma_1{}^2/\sigma_2{}^2) \cdot (u_2/u_1)^2\,[1 + \sigma_2{}^2/u_2{}^2 + \sigma_1{}^2/u_1{}^2] \qquad (5)$$

The asymptotic distribution of $I_{E/C}$ is

$$\sqrt{n}\,[I_{E/C} - (\sigma_1/u_1)/(\sigma_2/u_2)] \sim N(0, A\Sigma A') \qquad (6)$$

and the corresponding large sample normal approximation is

$$I_{E/C} \sim N[(\sigma_1/u_1)/(\sigma_2/u_2),\ A\Sigma A']. \qquad (7)$$

Substituting the respective sample statistics results in a test statistic of no difference between the treatment groups of

$$\frac{I_{E/C} - 1}{\sqrt{\sigma._{}^2/n}} \qquad (8)$$

and a corresponding confidence interval of

$$I_{E/C} - z_{\alpha/2}\,(\hat{\sigma}./\sqrt{n}) \leq CV_E/CV_C \leq I_{E/C} + z_{\alpha/2}\,(\hat{\sigma}./\sqrt{n}). \qquad (9)$$

For unequal sample sizes $\sigma.^2$ becomes

$$\sigma.^2 = [(\sigma_1/\sigma_2)\cdot(u_2/u_1)]^2 \left\{ \frac{N}{2}\cdot\frac{n_1 + n_2}{n_1\, n_2} + \frac{N}{n_2}(\sigma_2/u_2)^2 + \frac{N}{n_1}(\sigma_1/u_1)^2 \right\} \qquad (10)$$

where $N = n_1 + n_2$, and $I_{E/C} \sim N\,[(\sigma_1/\sigma_2)/(u_2/u_1),\ \sigma.^2/N]$.

ACCURACY OF THE LARGE SAMPLE APPROXIMATION

The statistical manipulations above result in a large sample approximation to the distribution of $I_{E/C}$. In order to check the usefulness of this approximation for small samples, a simulation study was conducted. The subroutine GGNML of the International Mathematical and Statistical Libraries (1977) was used to generate standard normal deviates. Table B1 gives the represenative values of $I_{E/C}$ used in the simulations. These were chosen to represent a range of values that would be considered likely in experimental studies such as the mastery learning studies under consideration here. Caution is recommended, however, in making interpretations beyond the range reported.

In each simulation, the experimental and control group sample sizes were chosen to be equal, that is $n = n_E = n_C$. The normal deviates were generated as two independent vectors of length n, one for the experimental group and one for the control group. The sample means and standard deviations for each were then found, allowing the computation of $I_{E/C}$. The means and variances of the simulated $I_{E/C}$'s, for the 14 types of simulation chosen, are given in Tables B2, B3 and B4. Also shown are the proportions of confidence intervals which contained the true value of $I_{E/C}$ with certain nominal significance levels. These proportions were found by noting, for each simulation, whether the true value of $I_{E/C}$ was within the nominal confidence interval calculated using the sample value of $I_{E/C}$ and the sample value of σ. (as given in equation 9); these instances were accumulated for each of the nominal significance levels and then divided by the total number of simulations to give the result. The proximity of thse proportions to the nominal level is an indication of the accuracy of the large sample approximation.

The first series of simulations are reported in Table B2; this series contrasts the performance of the approximation as the experimental standard deviation inflates and deflates, while the other parameters remain constant. Consider first the column of means; this column revealed a tendency to overestimate $I_{E/C}$. The bias declines from about 6 percent for the case n = 10, to about 2.5 percent for n = 20, down to about 1 percent for n = 50. This bias can be traced to the fundamental assymmetry in the distribution of the standard errors; it is of noteworthy size for the simulations at n = 10, and such cases should thereby be treated with caution, but its performance for the large sample sizes is quite within the bounds suitable for practical application. The story told by the proportion of confidence intervals containing $I_{E/C}$ is rather similar. Here, all values found match the nominal ones quite well except for the n = 10 cases. There is a small but marked tendency for the empirical proportion to be below the nominal ones. This may be due to bias of the estimator, a tendency for the variance estimator to underestimate the twice variation, or a combination of the two. This issue cannot be fully clarified until an unbiased estimator is found.

The second series of simulations is reported in Table B3; this series contrasts the performance of the approximation as the experimental mean shifts and the other parameters remain constant. The results give the same picture: the approximation is doing quite well for the case n = 20 and upwards, but is poor for n = 10. The third series of simulations is reported in Table B4; here the experimental means and standard deviations are simultaneously manipulated, but the experimental coefficient of variation and the control group parameters are kept constant. In this series also, the approximation holds up well for n = 20 and upward, but is poor for n = 10.

These results indicate that (subject to the previous remark on caution in extrapolating simulation results) the approximation is useful when there are 20 or more cases in each of the experimental and control groups, but that it should not be used for smaller sample groups. Work is continuing on finding a correction for the bias in the estimator $I_{E/C}$.

TABLE B1. Representative Values of $I_{E/C}$

Simulation	$I_{E/C}$	C	C	E	E
1	1.00	10.0	1.0	10.0	1.00
2	1.25	10.0	1.0	10.0	1.25
3	1.50	10.0	1.0	10.0	1.50
4	0.75	10.0	1.0	10.0	0.75
5	0.50	10.0	1.0	10.0	0.50
6	0.25	10.0	1.0	10.0	0.25
7	0.9091	10.0	1.0	11.0	1.00
8	0.8333	10.0	1.0	12.0	1.00
9	1.1111	10.0	1.0	9.0	1.00
10	1.25	10.0	1.0	8.0	1.00
11	1.00	10.0	1.0	14.0	1.40
12	1.00	10.0	1.0	12.0	1.20
13	1.00	10.0	1.0	8.0	0.80
14	1.00	10.0	1.0	6.0	0.60

TABLE B1. Representative Values of $I_{E/C}$

Simulation	$I_{E/C}$	C	C	E	E
1	1.00	10.0	1.0	10.0	1.00
2	1.25	10.0	1.0	10.0	1.25
3	1.50	10.0	1.0	10.0	1.50
4	0.75	10.0	1.0	10.0	0.75
5	0.50	10.0	1.0	10.0	0.50
6	0.25	10.0	1.0	10.0	0.25
7	0.9091	10.0	1.0	11.0	1.00
8	0.8333	10.0	1.0	12.0	1.00
9	1.1111	10.0	1.0	9.0	1.00
10	1.25	10.0	1.0	8.0	1.00
11	1.00	10.0	1.0	14.0	1.40
12	1.00	10.0	1.0	12.0	1.20
13	1.00	10.0	1.0	8.0	0.80
14	1.00	10.0	1.0	6.0	0.60

TABLE B3. Small Sample Accuracy of Confidence Intervals for $I_{E/C}$: Equal Standard Deviations and Differing Means

Sample Size $n = n_E = n_C$	Mean $I_{E/C}$	Variance $I_{E/C}$	Proportion of confidence intervals containing $I_{E/C}$ with nominal significance level .60	.70	.80	.90	.95	.99
		Simulation 7 $I_{E/C} = 0.9091$						
10	.9553	.1202	.576	.672	.763	.865	.910	.952
20	.9386	.0482	.599	.695	.785	.890	.947	.980
30	.9193	.0302	.602	.688	.789	.898	.942	.978
40	.9250	.0215	.608	.707	.807	.904	.952	.984
50	.9139	.0184	.583	.679	.790	.885	.941	.984
		Simulation 8 $I_{E/C} = 0.8333$						
10	.8958	.1047	.571	.680	.784	.873	.918	.962
20	.8564	.0421	.582	.686	.782	.880	.934	.969
30	.8452	.0271	.581	.686	.799	.886	.929	.971
40	.8467	.0204	.574	.674	.776	.889	.941	.978
50	.8383	.0150	.606	.700	.798	.886	.938	.985
		Simulation 9 $I_{E/C} = 1.1111$						
10	1.1865	.1849	.563	.667	.775	.870	.914	.952
20	1.1483	.0792	.596	.695	.794	.887	.936	.978
30	1.1293	.0479	.579	.690	.785	.886	.936	.979
40	1.1267	.0337	.603	.702	.802	.894	.944	.983
50	1.1216	.0290	.586	.687	.790	.881	.929	.980
		Simulation 10 $I_{E/C} = 1.25$						
10	1.3273	.2561	.558	.665	.771	.869	.920	.962
20	1.2830	.0986	.581	.689	.791	.887	.936	.976
30	1.2704	.0609	.588	.676	.786	.893	.940	.976
40	1.2771	.0469	.580	.676	.783	.891	.945	.985
50	1.2642	.0355	.565	.666	.778	.880	.941	.981

NOTE: Each of the simulations is based on 2000 cases.

TABLE B4. Small Sample Accuracy of Confidence Intervals for $I_{E/C}$: Differing Means and Standard Deviations

Sample Size $n = n_E = n_C$	Mean $I_{E/C}$	Variance $I_{E/C}$	Proportion of confidence intervals containing $I_{E/C}$ with nominal significance level .60	.70	.80	.90	.95	.99
		Simulation 11	$I_{E/C} = 1.0$					
10	1.0620	.1593	.552	.645	.745	.846	.899	.948
20	1.0224	.0649	.579	.678	.782	.872	.917	.972
30	1.0097	.0364	.596	.693	.792	.893	.944	.979
40	1.0159	.0285	.600	.694	.791	.900	.946	.986
50	1.0066	.0212	.595	.694	.793	.897	.945	.981
		Simulation 12	$I_{E/C} = 1.0$					
10	1.0850	.1712	.549	.652	.759	.868	.916	.955
20	1.0372	.0612	.585	.684	.789	.890	.941	.975
30	1.0180	.0407	.561	.664	.775	.876	.930	.979
40	1.0187	.0279	.591	.692	.798	.898	.945	.982
50	1.0081	.0227	.584	.593	.795	.892	.943	.985
		Simulation 13	$I_{E/C} = 1.0$					
10	1.0732	.1649	.572	.673	.774	.875	.918	.960
20	1.0328	.0653	.576	.678	.779	.876	.922	.966
30	1.0208	.0382	.601	.693	.795	.891	.939	.979
40	1.0103	.0285	.596	.689	.791	.885	.933	.980
50	1.0138	.0227	.595	.695	.796	.890	.941	.980
		Simulation 14	$I_{E/C} = 1.0$					
10	1.0676	.1579	.566	.666	.776	.880	.921	.959
20	1.0252	.0615	.585	.689	.784	.880	.931	.968
30	1.0158	.0382	.590	.694	.801	.887	.939	.977
40	1.0131	.0280	.598	.697	.797	.896	.938	.980
50	1.0131	.0217	.585	.685	.802	.902	.948	.983

NOTE: Each of the simulations is based on 2000 cases.