DOCUMENT RESUME

ED 440 978 TM 030 764

AUTHOR	Barnette, J. Jackson; McLean, James E.
TITLE	Empirically Based Criteria for Determining Meaningful Effect Size.
PUB DATE	1999-11-19
NOTE	37p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (28th, Point Clear, AL, November 17-19, 1999).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Criteria; *Effect Size; Monte Carlo Methods; *Prediction; Sample Size

ABSTRACT

The purpose of this study was to determine: (1) the extent to which effect sizes vary by chance; (2) the proportion of standardized effect sizes that achieve or exceed commonly used criteria for small, medium, and large effect sizes; (3) whether standardized effect sizes are random or systematic across numbers of groups and sample sizes; and (4) whether it is possible to predict standardized effect sizes using degrees of freedom, number of groups, and sample sizes. Monte Carlo procedures were used to generate standardized effect sizes in a one-way analysis of variance situation with 2 through 10 groups with samples sizes from 5 to 100 in steps of 5. Within each of the 180 configurations, 5,000 replications were done. It was found that standardized effect size variation was systematic rather than random. Numbers of groups and sample sizes were highly predictive of standardized effect size, but error degrees of freedom was not predictive. Equations were developed that could be used to predict standardized effect sizes that could be expected by chance, using number of groups and sample size as the predictor variables. The prediction equations were extremely accurate. This research provides a better alternative for the evaluation of empirical standardized effect sizes than the somewhat arbitrary and fixed criteria often used to classify standardized effect sizes as small, medium, or large. (Contains 3 tables, 10 figures, and 34 references.) (SLD)

Empirically Based Criteria for Determining Meaningful Effect Size

J. Jackson Barnette
University of Iowa

and

James E. McLean
University of Alabama at Birmingham

A Paper

Presented at the 1999 Annual Meeting
of the
Mid-South Educational Research Association
Point Clear, Alabama
November 19, 1999

For further information, contact:

Dr. Jack Barnette
College of Public Health
2811 Steindler Bldg.
University of Iowa
Iowa City, IA 52242
(319) 335 8905
jack-barnette@uiowa.edu

Abstract

The concept of effect size has become very important in educational research. Some have even advocated using effect size estimates in place of tests of statistical significance. Cohen's popular book titled *Statistical Power Analysis for the Behavioral Sciences* recommends specific levels of effect size for "small," "medium," and "large" effects. However, even Cohen acknowledges these values are relative to the specific content and method in a given research situation. The purpose of this study is to determine to what extent effect sizes vary by chance, how these conform to Cohen's levels, and if this variation is by chance.

Monte Carlo procedures were used to generate standardized effect sizes in a one-way ANOVA situation with 2 through 10 groups having sample sizes from 5 to 100 in steps of 5. Within each of the 180 number of group and sample size configurations, 5000 replications were done, all generated from a distribution of normal deviates. The process was tested by generating a known normal distribution and comparing it to its known characteristics.

It was found that standardized effect size variation was systematic rather than random. Number of groups and sample sizes were highly predictive of standardized effect size, but error degrees of freedom was not predictive. Equations were developed which could be used to predict standardized effect sizes that could be expected by chance, using number of groups and sample sizes as the predictor variables. The prediction equations were extremely accurate (R^2= 0.9990). Thus, this research provides a better alternative for the evaluation of empirical standardized effect sizes than the somewhat arbitrary and fixed criteria often used to classify standardized effect sizes as small, medium, or large.

Empirically Based Criteria for Determining Meaningful Effect Size

The concept of effect size has become very important in educational research. Some have even advocated using effect size estimates in place of tests of statistical significance (e.g., Carver, 1993; Nix & Barnette, 1998; Schmidt, 1996). Cohen's popular book titled *Statistical Power Analysis for the Behavioral Sciences* (1969, 1988) recommends specific levels of effect size for "small," "medium," and "large" effects. However, even Cohen acknowledges these values are relative to the specific content and method in a given research situation. The purpose of this study is to determine to what extent effect sizes vary by chance, how these conform to Cohen's levels, and if this variation is by chance.

The study used Monte Carlo methodology to address the following research questions:

1. To what extent do standardized effect sizes vary by chance?
2. What proportion of standardized effect sizes achieve or exceed commonly used criteria for small, medium, and large effect sizes?
3. Are standardized effect sizes random or systematic across number of groups and/or sample sizes?
4. Is it possible to reasonably predict standardized effect sizes, which would be expected by chance, using degrees of freedom, number of groups, and/or sample sizes?

The study was limited to the oneway analysis of variance situation with equal sample sizes. The number of groups ranged from 2 to 10 with each group having sample sizes ranging from 5 to 100 in steps of 5. Data were generated from normal deviates.

Background

The concept of effect size has been around for many years. Cohen (1969) is generally credited with coining the term. However, the development of meta-analysis by Glass, Rosenthal and others in the 1970s (e.g., Glass, 1976; 1978; Glass & Hakstian, 1969; Rosenthal, 1976, 1978) and the popularity of a book on meta-analysis in 1981 (Glass, McGaw, & Smith) are the catalysts for the interest in the concept. Numerous publications followed on applications of effect size methodology (e.g., Lynch, 1987; McLean, 1983), methods for estimating effect size and its properties (e.g., Fowler, 1988; 1993; Gibbons, Hedeker, & Davis, 1993;Hedges, 1981, 1984; Huynh, 1989; Kraemer, 1983; Reichhardt & Gollob, 1987; Thomas, 1986), extracting effect size estimates from existing studies (e.g., Hedges, 1982; Snyder & Lawson, 1993), and correcting effect size estimates (Snyder & Lawson, 1993). Another book by Wolf (1986) presented a general methodology for conducting meta-analysis including the extraction and testing of effect sizes.

Perhaps no one has had a greater impact on the use of effect sizes than Cohen (1977, 1988) through his books on power analysis. In these books, Cohen suggests general guidelines for levels of effect size. These are .2 for small effect, .5 for medium effect, and .8 for large effect. However, even Cohen was concerned about proposing these as standards. He stated:

> The terms "small," "medium," and "large" are relative, not only to each other, but to the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation. In the face of this relativity, there is a certain risk inherent in offering conventional operational definitions for these terms for use in power analysis in as diverse a field of inquiry as behavioral science. This risk is nevertheless accepted in the belief that more is to be gained than lost by supplying a common conventional frame of reference which is recommended for use only when no better basis for estimating the ES index is available. (1988, p. 25)

Cohen's concerns were cited by Wolf (1986) and suggests that effect sizes should be interpreted in context. Specifically, one possibility is to compare a given effect size to the median effect size of studies extracted from the professional literature in that specific context rather than use some arbitrary guideline. Wolf indicates that a .5 standard deviation improvement is often considered practically significant and that the general guidelines of the National Institute of Education's Joint Dissemination Review Panel require .33 effect size, but at times will accept .25 to establish educational significance.

A broader debate on the use of statistical significance testing emerged from Cohen's power analysis books and other works. Kaufman (1998) indicates that the "controversy about the use or misuse of statistical significance testing has been evident in the literature for the past 10 years and has become the major methodological issue of our generation" (p. 1). The debate has spawned at least two special issues of journals (*Research in the Schools*, McLean & Kaufman, 1998; *Journal of Experimental Education*, Thompson, 1993) and dozens of other articles. The editorial policies of journals have been changed by the debate (e.g., APA, 1994; Schafer, 1990, 1991; Thompson, 1994, 1997).

The debate has ranged from those who recommend the elimination of statistical significance testing (e.g., Carver, 1978, 1993; Nix & Barnette, 1998) to those who staunchly support it (e.g., Frick, 1996; Levin, 1993, 1998; McLean & Ernest, 1998). However, even those who defend statistical significance testing indicate that significant results should be accompanied by a measure of practical significance. The leading method of reporting practical significance is through the provision of an effect size estimate (Kirk, 1996; McLean & Ernest, 1998; Robinson & Levin, 1997; Thompson, 1996). Unfortunately, the criteria for judging the practical significance of results based on effect size has defaulted to the use of Cohen's (1988) guidelines that even Cohen has warned us about (1977, 1988, 1990). As Wolf (1986) noted, empirical standards for judging effect size are needed.

Methodology

Monte Carlo methods were used to generate the data for this research. All data were generated from a random normal deviate routine, which was incorporated into a larger compiled QBASIC program. All sampling and computation, conducted with double-precision, routines were verified using SAS® programs. The program was run on a Dell Pentium II, 266 MHz personal computer. Final analysis of the standardized effect sizes was conducted using SAS® and Microsoft Excel.

Some preliminary analyses were run using the Monte Carlo program to test its accuracy. First, 500,000 standard normal scores (z-scores) were generated and the statistics for the distribution were computed. This resulted in a mean = -.00096, variance = 1.0013, skewness =

.00056, kurtosis = .00067, and the Wilk-Shapiro D = .000734 (nonsignificant). Thus, we concluded that the program generates reasonable normal distributions. Second, 900,000 cases were computed with K ranging from 2 to 10 and n ranging from 5 to 100 with no differences between the group means. In each case, the proportions of significant F-statistics were computed corresponding to preset alphas of .25, .10, .05, .01, .001, and .0001. The resulting proportions of rejected null hypotheses were .24989, .10106, .05071, .01022, .001004, and .000103 respectively. These results support the accuracy of the Monte Carlo program.

Standardized effect sizes were generated for 5,000 replications within each combination of number of groups from 2 to 10 and sample sizes from 5 to 100 in steps of 5. The standardized effect size was computed as the range of means divided by the root mean square error. Within each number of group and sample size configuration several statistics were determined including: range, mean, and variance of SES values and proportions of observed SES values that achieved or exceeded the Cohen proposed criteria for small, medium, and large SES values. In addition, a data file was created which included (for each number of groups and sample size configuration) number of groups, sample size, and mean SES. These data were used to generate total and error degrees of freedom values. Analysis of data in this file included the use of SAS® for summary statistics and the trendline analysis program from Excel.

Results

Table 1 presents the mean standardized effect sizes for selected number of groups and sample size configurations. While number of groups ranged from 2 to 10, only K=2, 3, 4, 6, 8, and 10 are reported and sample sizes ranged from 5 to 100, in units of 5, but sample sizes of 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, and 100 are reported. Marginal totals included all data, not just data reported in the individual table cells. The mean standardized effect size for the 180 configurations was 0.4065 with a range of 0 to +4.339 and standard deviation of 0.2927. Means and standard deviations are presented for selected K and n totals. Figure 1 presents the results for mean SES by number of groups with different patterns from low to high representing larger sample sizes to smaller sample sizes. Figure 2 presents the mean SES values, collapsed across sample sizes, for each number of groups along with +/- 1 standard deviation bars. It is clear that as number of groups increases so does mean SES.

Figure 3 presents the results for SES by sample sizes with different patterns from low to high representing larger numbers of groups to smaller numbers of groups. Figure 4 presents the mean SES values for each sample size along with +/- 1 standard deviation bars, collapsed across number of groups. It is clear that as sample size increases, SES decreases. It is very apparent that mean, chance dependent, SES values are affected by both number of samples and sample size. It is reasonable to expect that since both of these factors relate to error degrees of freedom, this might have a direct influence on SES. However, it is clear from examination of Figure 5 that error degrees of freedom does not provide a systematic and unequivocal function which could be used to predict SES.

Table 2 presents the proportions of observed mean SES values equal to or exceeding the Cohen proposed criteria for small, medium, and large effect sizes. More than 80% of the observed SES values achieved the small effect size criterion, almost 25% achieved the medium effect size criterion, and more than 8% achieved the large effect size criterion, all as a function of chance. Clearly the combinations of larger numbers of groups combined with smaller sample sizes had the highest proportions of achieving the criteria while smaller numbers of groups combined with larger sample sizes were less likely to have chance generated SES values achieving the criteria standards.

Examination of the relationship patterns between mean SES and sample size within each of the number of group situations, as presented in Table 3, indicates that every one of them followed a power function, with coefficients determined by the trend-line function of Microsoft Excel. These coefficients were labeled as "a" and "b" in the power function of $M_{ses} = an^{-b}$, where n is the sample size. The next step was to determine if these coefficients could be found to be functions of K, the number of samples. Factor a was a logarithmic function of K, as a=

1.1498Ln(K) + 0.5374, with an R^2 of 0.9965. While it was not as strong a relationship (R^2= 0.9290), one that may be improved with further analysis, there was a quadratic relationship of factor b as related to K. Factor b, as related to K, was determined to be b= $0.0006K^2 - 0.009K + 0.5411$. These then became the functions of K to use in the prediction of mean chance-determined, SES based on K and sample size.

This equation was used to predict SES values for the 180 number of groups, sample size configurations. The relationship between the observed SES and the SES predicted using the empirically-determined equation is presented in Figure 6. The prediction was almost exact, having an R^2 of 0.9990. While it may be possible to further refine the coefficients, this prediction of SES, by chance, based on K and n is very useable and very accurate.

Table 3 presents the coefficients that would be used for each level of the "number of groups" variable to predict mean SES based on sample size for number of groups from 2 through 10 and for sample sizes of 5 to 100. Figures 7 through 10 graphically display the relationship between sample size and mean SES for K= 2, 3, 6, and 10 respectively, based on the equations found in Table 3. The equations, or graphic representations, could be used to predict the mean SES one would expect to get by chance for any sample size from 5 to 100 in a given number of groups of two through ten condition.

Conclusions

To what extent do standardized effect sizes vary by chance? Standardized effect sizes vary greatly by chance. The largest SES was 4.339, the mean was 0.4065 and the standard deviation was 0.2927. In the two-sample (t test) situation the SES ranged from 0 to 3.812 with a mean of 0.1972. Within the K= 2 situation, the largest SES was found in the smallest (n= 5) sample, a mean SES of 0.5601. In the largest number of groups (K= 10), the mean SES was 0.5253, with a range of 0.077 to 3.312. Clearly, standardized effect sizes do vary by chance.

What proportion of standardized effect sizes achieve or exceed commonly used criteria for small, medium, and large effect sizes? A very high proportion of the mean SES's (0.8040) meet or exceed the 0.20 small effect size criterion, about a fourth (0.2453) meet or exceed the 0.50 medium effects size criterion, and 0.0837 meet or exceed the large effect size criterion. Thus, a very high proportion of mean SES values meet or exceed the commonly used (Cohen) criteria labeled as small, medium, or large effect sizes by chance.

Are standardized effect sizes random or systematic across number of groups and or sample sizes? Effect size differences are clearly not random across numbers of groups or sample sizes. Mean SES values increase as number of groups increase and decrease as sample sizes increase in systematic patterns.

Is it possible to reasonably predict standardized effect sizes that would be expected by chance using error degrees of freedom, number of groups, and/or sample sizes? Degrees of freedom error does not provide for systematic prediction of mean SES. The number of groups (K) and the sample size (n) are systematically predictive of mean SES. An initial, empirically-derived, equation that can be used to make reasonable prediction of mean SES as a function of k and n is:

$M_{ses} = a\, n^{-b}$

Where $a = 1.1498\ Ln(K) + 0.5374$ and $b = 0.0006K^2 - 0.009K + 0.5411$

When this equation is used to predict the 180 observed mean SES values generated by the Monte Carlo program, the R^2 for the relationship of predicted and observed values is 0.9990. While it may be possible to improve the accuracy of the prediction equation slightly, this equation, or graphic generations using these equations could be used with confidence to predict expected values of mean standardized effect size in any situation of two to ten groups with equal sample sizes of 5 to 100.

Is it more reasonable to compare observed standardized effect sizes with criteria such as those suggested by Cohen and others that are fixed and arbitrary, or ones predicted by the equations generated in this research? Clearly, many standardized effect sizes meet or exceed these values and the extent to which they do this is systematically related to number of groups and sample sizes. Our approach takes into account number of groups and sample sizes in predicting standardized effect sizes that would be obtained by chance. In practice this should be used to evaluate observed standardized effect sizes. It is possible to have a standardized effect size meeting the criteria of a "medium" effect size and even a "large" effect size that could be a chance event. Using our prediction equation allows for at least judging whether an observed standardized effect size is lower, about equal, or higher than one expected by chance in relation to number of groups and sample sizes. This, clearly, is a preferred approach. It is more realistic and accurate than the use of a fixed and arbitrary set of judgmental criteria.

Needed Research

While this study provides pretty convincing evidence that the use of Cohen's criteria (1988) for judging practical significance is risky, questions remain. This study was limited to equal sample sizes. In unequal sample size situations, what n should be used? Should we use the mean sample size or possibly the harmonic mean as is done with many multiple comparison procedures? In this study, 5,000 replications were completed for each combination of number of groups sample size. Would using more replicates result in a refinement of the coefficients? This study was also limited to oneway ANOVAs. What results might we get from multi-factor ANOVA situations?

Another area of research might be the examination of how these relationships are related using other measures of effect size, based on other statistics such as correlation coefficients and tests on proportions. In addition, this approach could be used to predict other measures of effect size, such as the effect size indices proposed by Cohen and measures of association such as eta-squared and omega-squared.

References

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review, 48*, 378-399.

Carver, R. P. (1993). The case against statistical significance testing, revisited. Journal of Experimental Education, 61(4), 287-292.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*(12), 1304-1312.

Fowler, R. L. (1988). Estimating the standardized mean difference in intervention studies. *Journal of Educational Statistics, 13*(4), 337-350.

Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods, 1*(4), 379-390.

Glass, G. V. (1978). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5*, 351-379.

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

Glass, G. V., & Hakstian, A. R. (1969). Measures of association in comparative experiments: Their development and interpretation. *American Educational Research Journal, 6*, 403-414.

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

Gibbons, R. D., Hedeker, D. R., & Davis, J. M. (1993). Estimation of effect sizes from a series of experiments involving paired comparisons. *Journal of Educational Statistics, 18*(3), 271-279.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 2*(2), 107-128.

Hedges, L. V. (1982). *Statistical methodology in meta-analysis*. (ERIC Document Reproduction Service No. ED 227 133).

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics, 9*(1), 61-85.

Huynh, C. -L. (1989). *A unified approach to the estimation of effect size in meta-analysis*. A paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 306 248).

Kaufman, A. S. (1998). Introduction to the special issue on statistical significance testing. *Research in the Schools*, 5(2), 1.

Kraemer, H. C. (1983). Theory of estimation and testing of effect sizes: Use in meta-analysis. *Journal of Educational Statistics, 8*(2), 93-101.

Lynch, K. B. (1987). The size of educational effects: An analysis of programs reviewed by the Joint Dissemination Review Panel. *Educational and Policy Analysis, 9*(1), 55-61.

McLean, J. E. (1983). *A meta-analysis approach to impact evaluation of adoptions*. Paper presented at the National Diffusion Network Regional Meeting, Memphis, TN (ERIC Document Reproduction Service No. ED 242 744).

McLean, J. E., & Ernest, J. M. (1998). The role of statistical significance testing in educational research. *Research in the Schools, 5*(2), 15-22.

McLean, J. E., & Kaufman, A. S. (Eds.). (1998). Statistical significance testing [Special Issue]. *Research in the Schools, 5*(2).

Nix, T. W., & Barnette, J. J. (1998). The data analysis dilemma: Ban or abandon. A review of null hypothesis significance testing. *Research in the Schools, 5*(2), 3-14.

Reichardt, C. S., & Gollob H. F. (1987). Taking uncertainty into account when estimating effects. *New Directions for Program Evaluation., 35*, 7-22.

Reynolds, S., & Day, J. (1984). Monte Carlo studies of effect size estimates and their approximations in meta-analysis. A paper presented at the annual meeting of the American Educational Research Association, Toronto, ON, Canada. (ERIC Document Reproduction Service No. ED 253 567).

Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York: Irvington.

Rosenthal, R. (1978). Combining the results of independent studies. *Psychological Bulletin, 85*, 185-193.

Schafer, W. D. (1990). Interpreting statistical significance. *Measurement and Evaluation in Counseling and Development, 23*, 98-99.

Schafer, W. D. (1991). Power analysis in interpreting statistical significance. *Measurement and Evaluation in Counseling and Development, 24, 146-148.*

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61*(4), 334-349.

Thomas, H. (1986). Effect size standard errors for the non-normal non-identically distributed case. *Journal of Educational Statistics, 11*(4), 293-303.

Thompson, B. (Guest Ed.). (1993). Statistical significance testing in contemporary practice [Special Issue]. *The Journal of Experimental Education, 61*(4).

Wolf, F. M. (1986). *Meta-Analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage Publications.

Table 1. Mean Standardized Effect Size by Number of Samples and Sample Size

n	Statistic	K= 2	K= 3	K= 4	K= 6	K= 8	K= 10	Total
5	M	0.5601	0.8272	0.9674	1.1825	1.3037	1.4044	1.1023
	SD	0.4640	0.4809	0.4617	0.4372	0.4135	0.3988	0.5092
	Min.-Max.	0.000-3.812	0.004-4.339	0.013-3.798	0.158-3.559	0.267-3.178	0.302-3.312	0.000-4.339
10	M	0.3781	0.5460	0.6628	0.8119	0.9152	0.9804	0.7626
	SD	0.3025	0.3015	0.2941	0.2878	0.2766	0.2678	0.3438
	Min.-Max.	0.000-2.167	0.011-2.073	0.014-1.952	0.118-2.500	0.167-2.344	0.278-2.157	0.000-2.500
15	M	0.2999	0.4421	0.5398	0.6617	0.7354	0.7996	0.6177
	SD	0.2341	0.2390	0.2398	0.2283	0.2182	0.2136	0.2754
	Min.-Max.	0.000-1.623	0.004-1.580	0.016-1.758	0.126-1.670	0.127-1.678	0.177-1.657	0.000-1.873
20	M	0.2595	0.3844	0.4640	0.5673	0.6376	0.6914	0.5340
	SD	0.1965	0.2080	0.2048	0.1949	0.1887	0.1834	0.2375
	Min.-Max.	0.000-1.292	0.003-1.659	0.023-1.416	0.071-1.381	0.147-1.377	0.206-1.472	0.000-1.659
25	M	0.2281	0.3419	0.4132	0.5134	0.5709	0.6200	0.4768
	SD	0.1724	0.1830	0.1779	0.1746	0.1664	0.1650	0.2109
	Min.-Max.	0.000-1.085	0.004-1.361	0.027-1.264	0.075-1.368	0.089-1.262	0.147-1.286	0.000-1.529
30	M	0.2106	0.3109	0.3826	0.4623	0.5194	0.5647	0.4344
	SD	0.1605	0.1693	0.1638	0.1560	0.1544	0.1480	0.1920
	Min.-Max.	0.000-1.070	0.003-1.172	0.010-1.166	0.055-1.205	0.115-1.367	0.158-1.289	0.000-1.367
40	M	0.1792	0.2709	0.3290	0.4017	0.4497	0.4869	0.3756
	SD	0.1374	0.1451	0.1415	0.1363	0.1302	0.1284	0.1659
	Min.-Max.	0.000-1.046	0.005-0.950	0.014-0.905	0.028-1.048	0.101-1.182	0.152-0.986	0.000-1.182
50	M	0.1621	0.2388	0.2900	0.3606	0.4042	0.4351	0.3361
	SD	0.1243	0.1268	0.1251	0.1209	0.1188	0.1132	0.1483
	Min.-Max.	0.000-0.762	0.001-0.780	0.019-0.868	0.057-0.980	0.116-0.958	0.121-0.913	0.000-1.084
60	M	0.1453	0.2195	0.2654	0.3262	0.3682	0.3976	0.3062
	SD	0.1104	0.1151	0.1165	0.1092	0.1056	0.1044	0.1352
	Min.-Max.	0.000-0.773	0.005-0.741	0.014-0.849	0.037-0.779	0.073-0.807	0.077-1.000	0.000-1.000
70	M	0.1341	0.2040	0.2471	0.3030	0.3426	0.3681	0.2840
	SD	0.1037	0.1081	0.1054	0.1018	0.0984	0.0964	0.1253
	Min.-Max.	0.000-0.575	0.004-0.808	0.015-0.738	0.026-0.808	0.081-0.789	0.124-0.770	0.000-0.808
80	M	0.1257	0.1922	0.2298	0.2832	0.3177	0.3454	0.2653
	SD	0.0951	0.1019	0.0997	0.0962	0.0918	0.0899	0.1168
	Min.-Max.	0.000-0.671	0.003-0.665	0.012-0.737	0.046-0.742	0.071-0.747	0.090-0.756	0.000-0.798
90	M	0.1195	0.1795	0.2175	0.2677	0.3000	0.3244	0.2504
	SD	0.1828	0.0950	0.0919	0.0898	0.0865	0.0860	0.1101
	Min.-Max.	0.000-0.588	0.002-0.636	0.016-0.586	0.034-0.648	0.089-0.735	0.084-0.668	0.000-0.735
100	M	0.1152	0.1707	0.2057	0.2542	0.2857	0.3084	0.2378
	SD	0.0884	0.0904	0.0877	0.0868	0.0822	0.0800	0.1047
	Min.-Max.	0.000-0.533	0.007-0.602	0.016-0.691	0.030-0.724	0.071-0.611	0.080-0.618	0.000-0.724
Total	M	0.1972	0.2935	0.3541	0.4346	0.4861	0.5253	0.4065
	SD	0.2070	0.2406	0.2558	0.2818	0.2964	0.3097	0.2927
	Min.-Max.	0.000-3.812	0.001-4.339	0.003-3.798	0.026-3.559	0.066-3.178	0.077-3.312	0.000-4.339

Note: Totals are based on K of 2 through 10 and n of 5 through 100 in steps of 5.

Table 2. Proportion of Mean Standardized Effect Sizes Achieving or Exceeding "Criterion" by Number of Samples and Sample Size

n	Effect Size	K= 2	K= 3	K= 4	K= 6	K= 8	K= 10	Total
5	Small, .20	0.7596	0.9478	0.9868	0.9996	1.0000	1.0000	0.9659
	Medium, .50	0.4594	0.7284	0.8532	0.9686	0.9936	0.9976	0.8794
	Large, .80	0.2440	0.4578	0.6016	0.8052	0.9056	0.9588	0.7205
10	Small, .20	0.6684	0.8956	0.9710	0.9980	0.9998	1.0000	0.9471
	Medium, .50	0.2810	0.4988	0.6840	0.8710	0.9532	0.9808	0.7727
	Large, .80	0.1000	0.1910	0.2908	0.4758	0.6366	0.7436	0.4528
15	Small, .20	0.5872	0.8452	0.9458	0.9946	0.9988	0.9998	0.9279
	Medium, .50	0.1874	0.3580	0.5324	0.7442	0.8694	0.9364	0.6661
	Large, .80	0.0388	0.0820	0.1410	0.2560	0.3538	0.4704	0.2520
20	Small, .20	0.5350	0.8018	0.9156	0.9882	0.9978	1.0000	0.9110
	Medium, .50	0.1268	0.2614	0.3948	0.6060	0.7582	0.8502	0.5570
	Large, .80	0.0142	0.0388	0.0648	0.1186	0.1872	0.2652	0.1307
25	Small, .20	0.4874	0.7544	0.8962	0.9812	0.9978	0.9996	0.8956
	Medium, .50	0.0750	0.1884	0.2902	0.5020	0.6434	0.7592	0.4573
	Large, .80	0.0050	0.0148	0.0264	0.0644	0.0930	0.1384	0.0633
30	Small, .20	0.4506	0.7138	0.8740	0.9688	0.9930	0.9984	0.8797
	Medium, .50	0.0624	0.1376	0.2226	0.3712	0.5214	0.6490	0.3668
	Large, .80	0.0036	0.0082	0.0122	0.0318	0.0468	0.0640	0.0300
40	Small, .20	0.3740	0.6412	0.8096	0.9452	0.9842	0.9970	0.8456
	Medium, .50	0.0300	0.0756	0.1264	0.2270	0.3256	0.4342	0.2278
	Large, .80	0.0008	0.0022	0.0018	0.0052	0.0114	0.0148	0.0063
50	Small, .20	0.3300	0.5758	0.7428	0.9222	0.9732	0.9930	0.8147
	Medium, .50	0.0136	0.0344	0.0626	0.1254	0.2096	0.2742	0.1350
	Large, .80	0.0000	0.0000	0.0004	0.0016	0.0020	0.0022	0.0012
60	Small, .20	0.2638	0.5216	0.6844	0.8848	0.9570	0.9834	0.7779
	Medium, .50	0.0080	0.0168	0.0342	0.0650	0.1106	0.1668	0.0757
	Large, .80	0.0000	0.0000	0.0006	0.0000	0.0002	0.0010	0.0003
70	Small, .20	0.2364	0.4678	0.6404	0.8446	0.9366	0.9762	0.7466
	Medium, .50	0.0046	0.0098	0.0138	0.0346	0.0658	0.0936	0.0422
	Large, .80	0.0000	0.0002	0.0000	0.0002	0.0000	0.0000	0.0000
80	Small, .20	0.2018	0.4212	0.5804	0.8012	0.9070	0.9622	0.7106
	Medium, .50	0.0020	0.0056	0.0098	0.0222	0.0308	0.0538	0.0230
	Large, .80	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
90	Small, .20	0.1828	0.3716	0.5406	0.7670	0.8802	0.9428	0.6773
	Medium, .50	0.0006	0.0032	0.0038	0.0104	0.0168	0.0330	0.0124
	Large, .80	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	Small, .20	0.1698	0.3418	0.4934	0.7212	0.8528	0.9212	0.6435
	Medium, .50	0.0002	0.0022	0.0032	0.0070	0.0102	0.0154	0.0072
	Large, .80	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Total	Small, .20	0.3560	0.5914	0.7389	0.8906	0.9522	0.9792	0.8040
	Medium, .50	0.0667	0.1265	0.1785	0.2615	0.3251	0.3794	0.2453
	Large, .80	0.0204	0.0400	0.0573	0.0887	0.1130	0.1348	0.0837

Note: Totals are based on K of 2 through 10 and n of 5 through 100 in steps of 5.

Table 3. Prediction Equations for Mean Standardized Effect Sizes of Form $M_{ses} = a\,n^{-b}$ by Number of Groups

K	Observed Equation Coefficients			Final Equation Coefficients	
	a	b	R^2	a	b
2	1.2727	0.5280	0.9990	1.3344	0.5255
3	1.8266	0.5171	0.9990	1.8006	0.5195
4	2.1631	0.5112	0.9997	2.1314	0.5147
5	2.4210	0.5092	0.9998	2.3879	0.5111
6	2.6443	0.5100	0.9998	2.5976	0.5087
7	2.7665	0.5052	0.9999	2.7748	0.5075
8	2.9154	0.5056	0.9999	2.9283	0.5075
9	3.0474	0.5062	1.0000	3.0638	0.5087
10	3.1473	0.5053	1.0000	3.1849	0.5111

Final equation coefficients as functions of K:

$$a = 1.1498\ \text{Ln}(K) + 0.5374$$

$$b = 0.0006K^2 - 0.009K + 0.5411$$

Figure 1. Standardized Effect Size by Number of Groups (K)

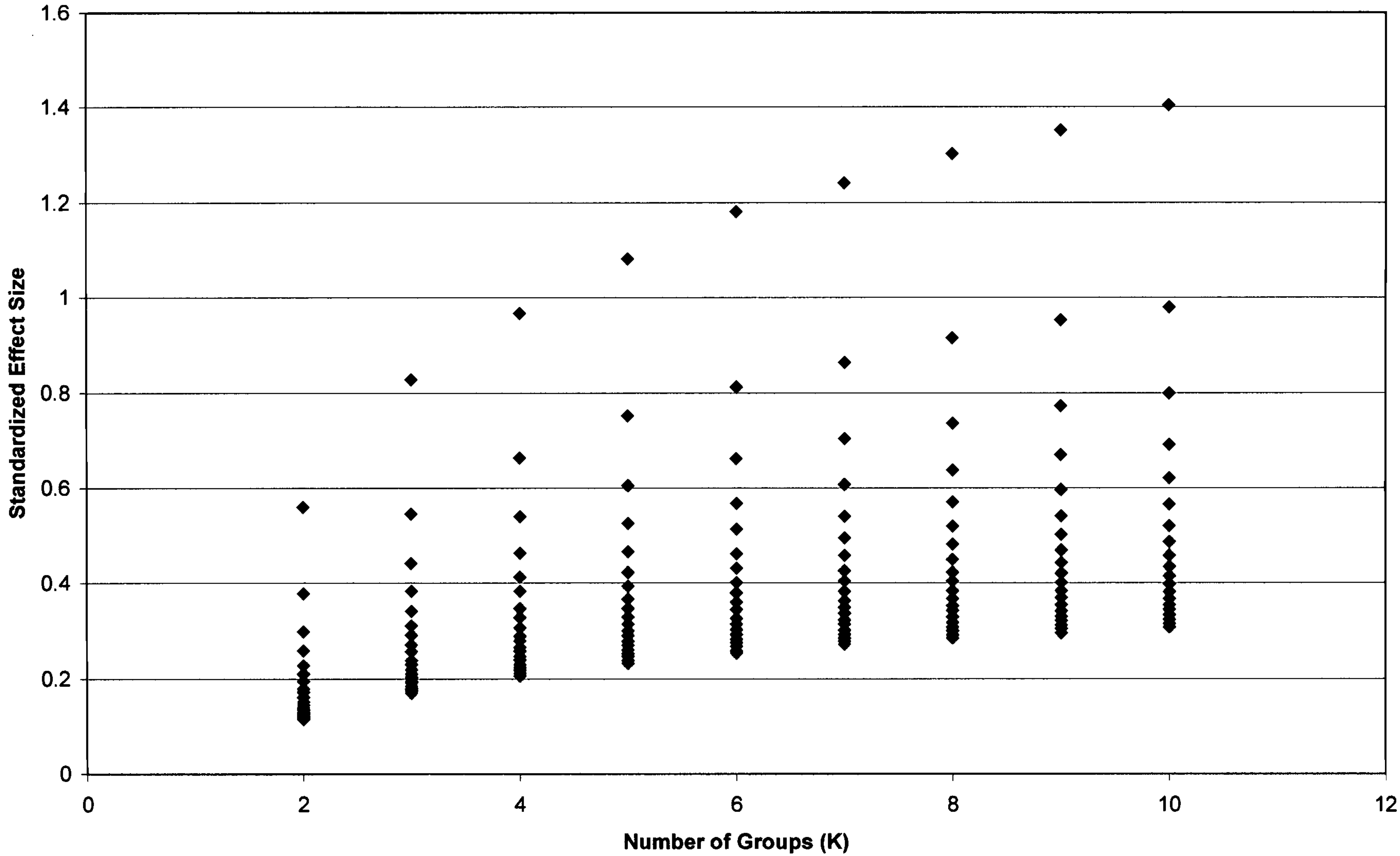


Figure 2. Standardized Effect Size by Number of Groups, Collapsed Across n +/- 1 Standard Deviation Bars

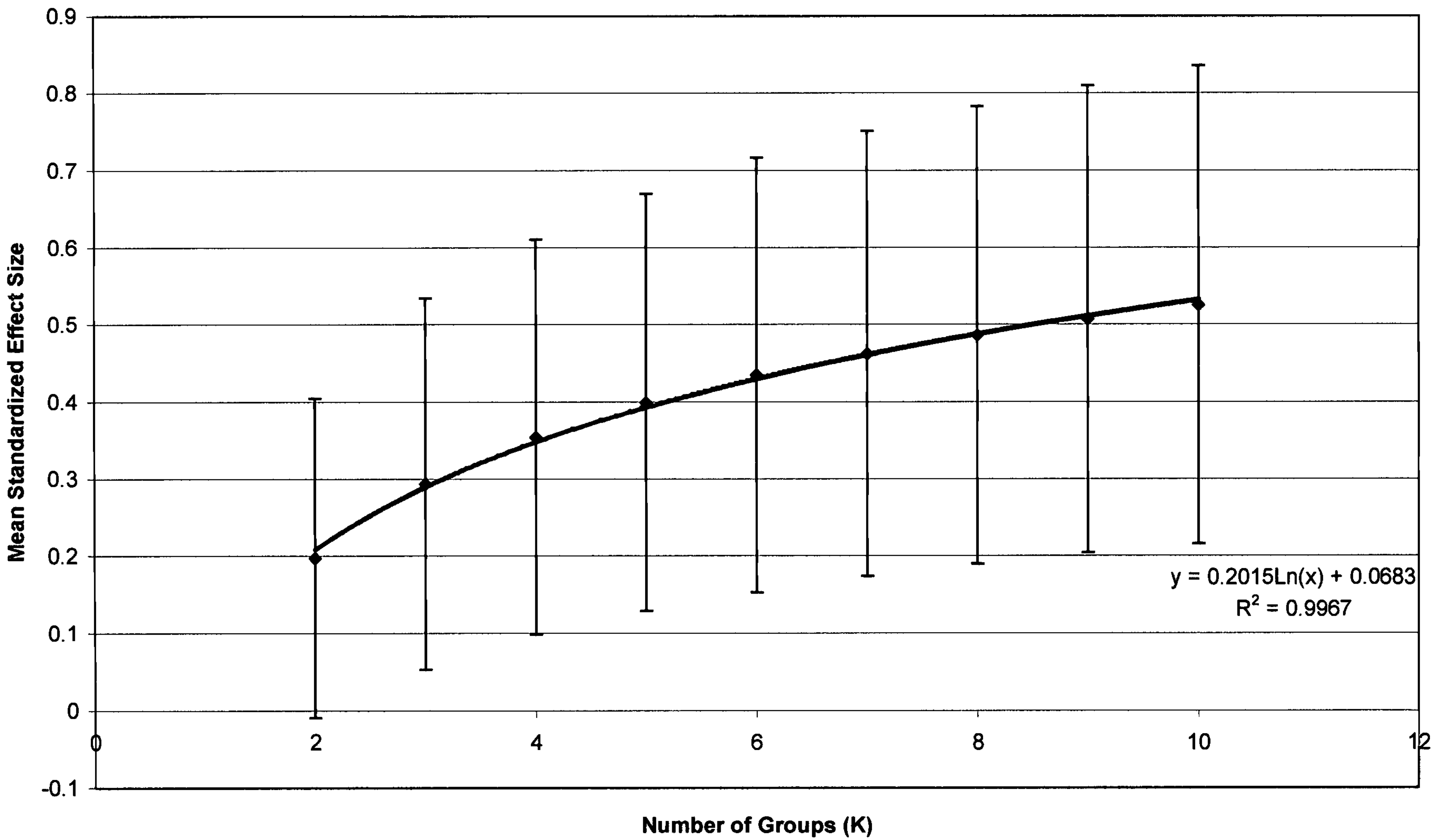


Figure 3. Standardized Effect Size by Sample Size (n)

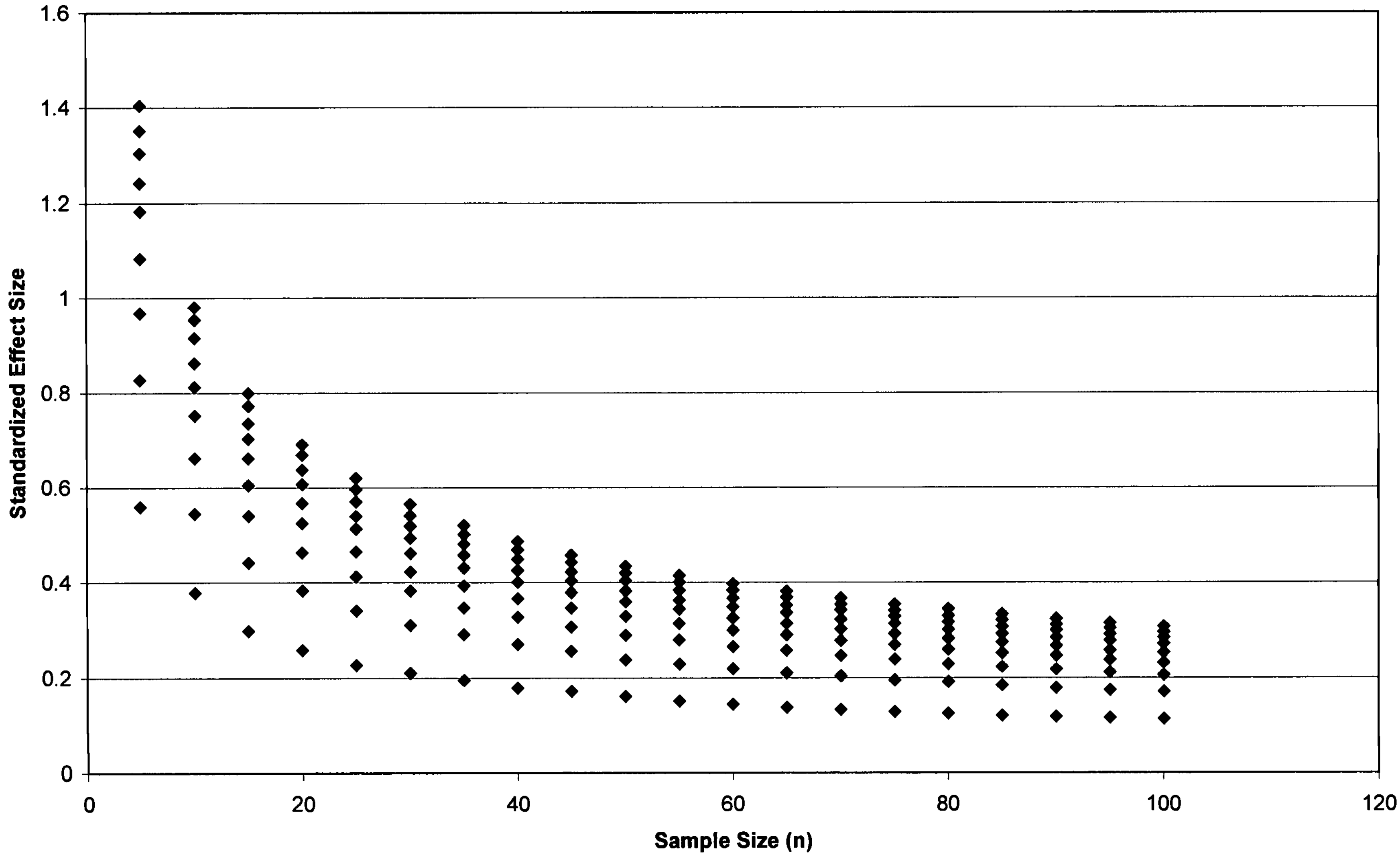


Figure 4. Standardized Effect Size by Sample Size, Collapsed Across K +/- 1 Standard Deviation Bars

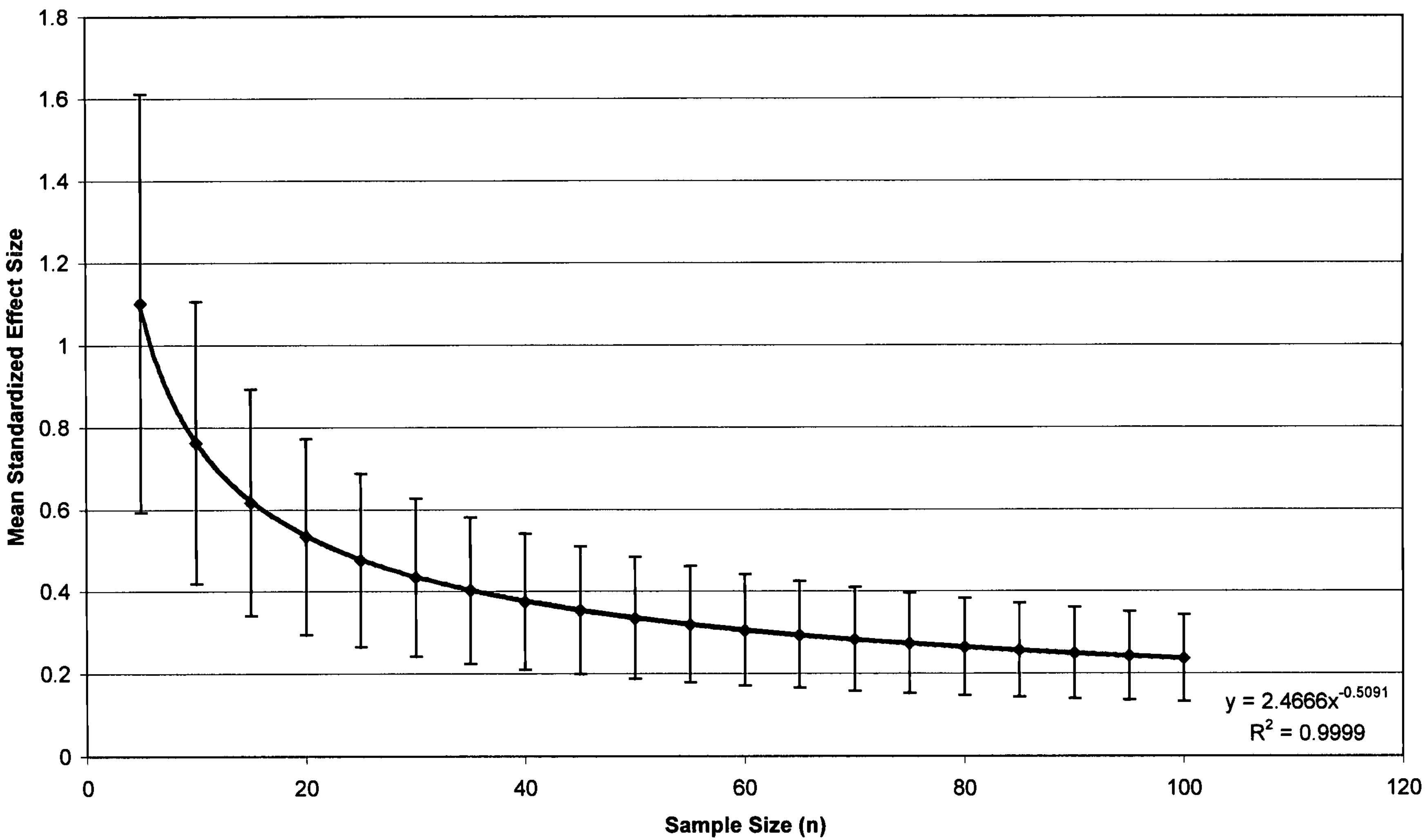


Figure 5. Relationship of Standardized Effect Size and Error Degrees of Freedom

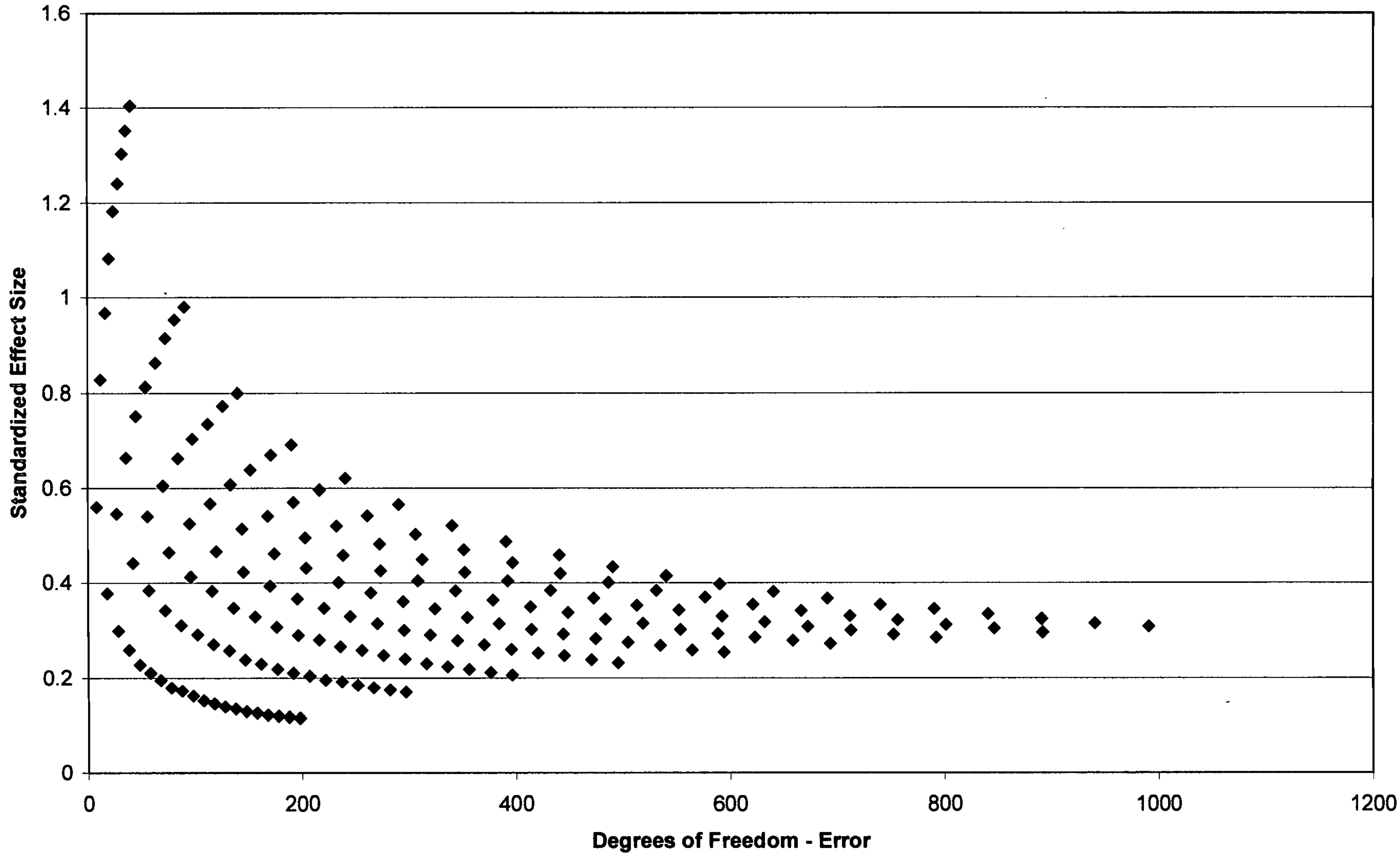


Figure 6. Relationship of Predicted and Observed Standardized Effect Sizes

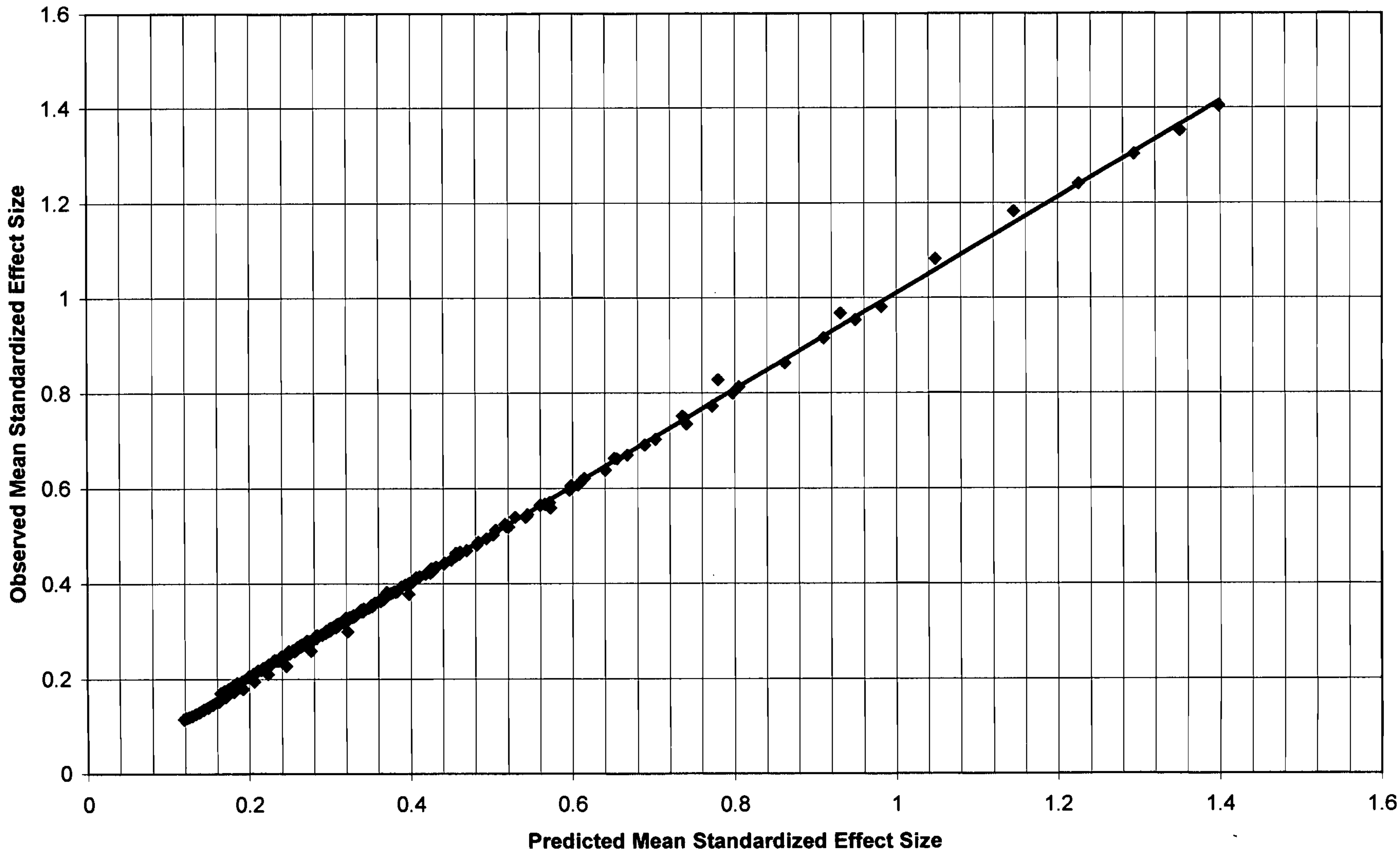
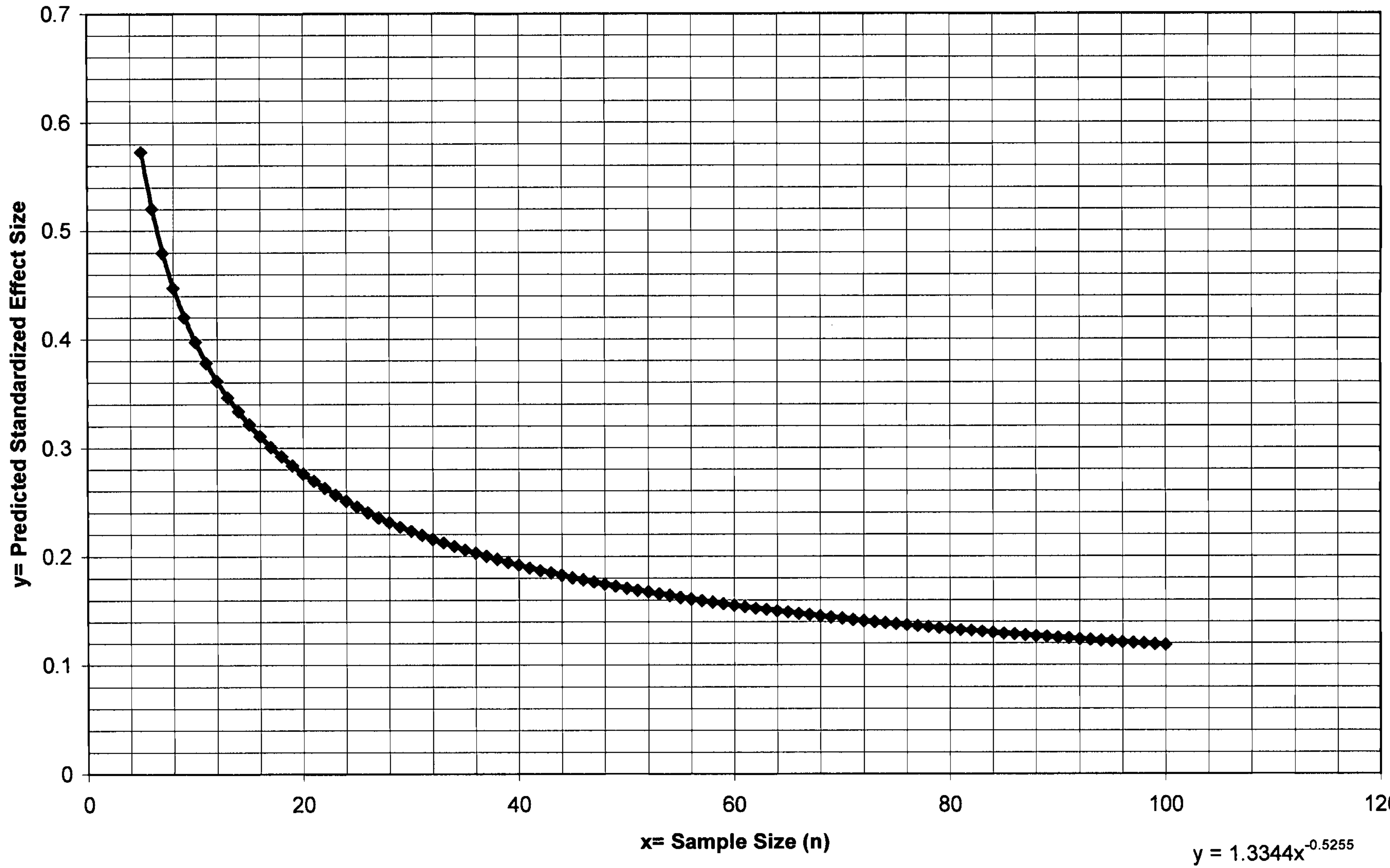


Figure 7. Predicted Standardized Effect Size by n for K= 2



30

31

Figure 8. Predicted Standard Effect Size by n for K= 3

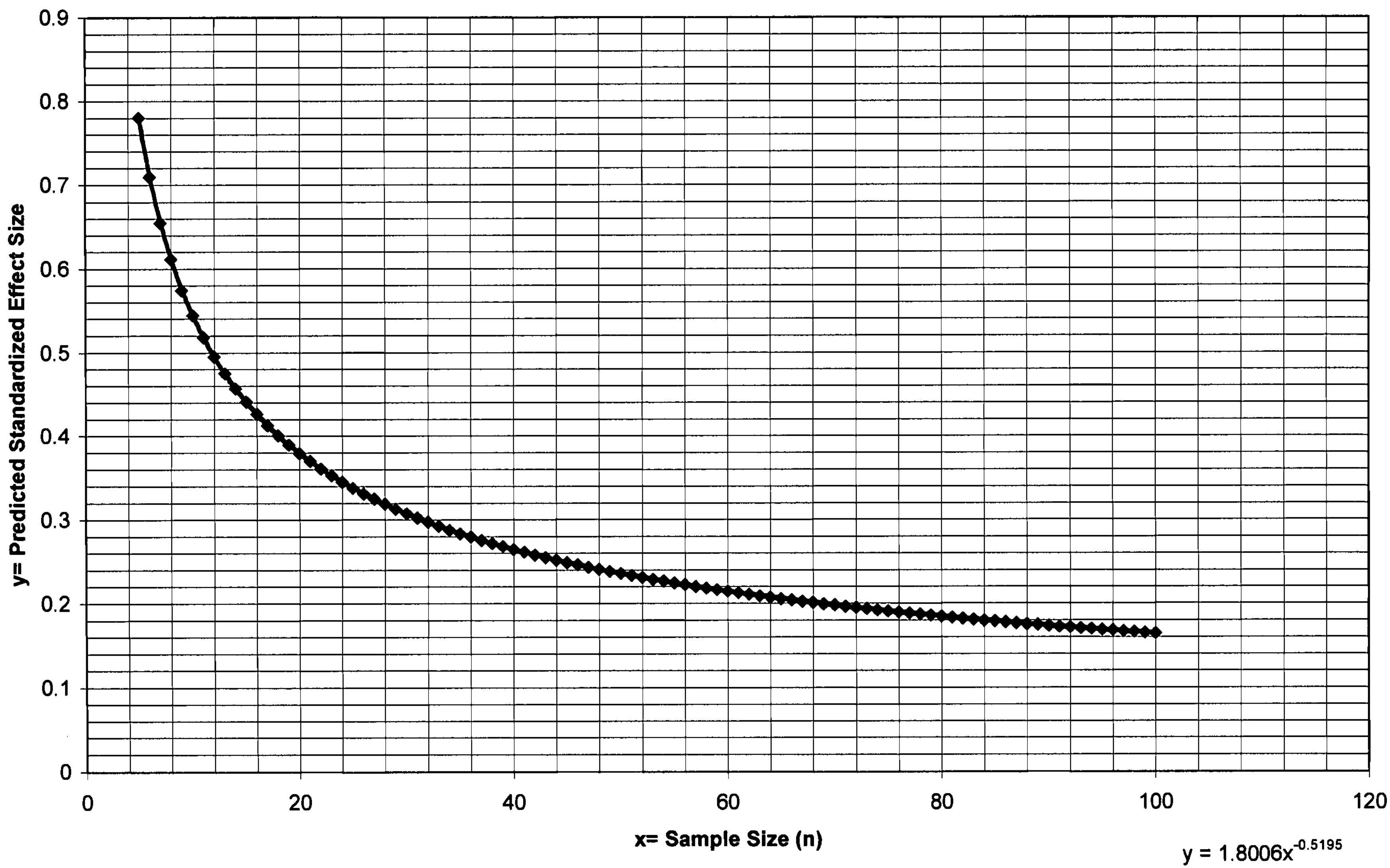


Figure 9. Predicted Standardized Effect Size by n for K= 6

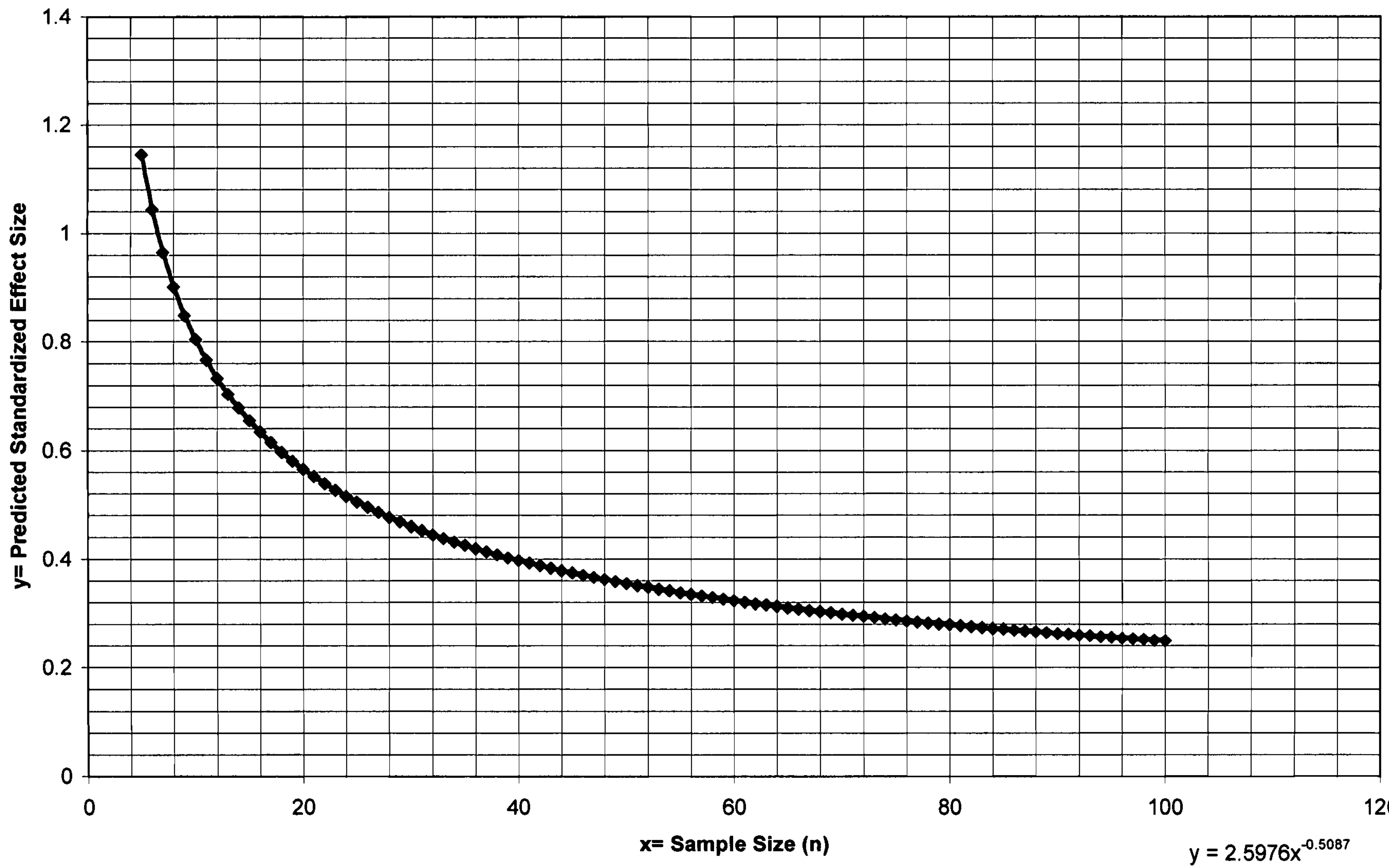
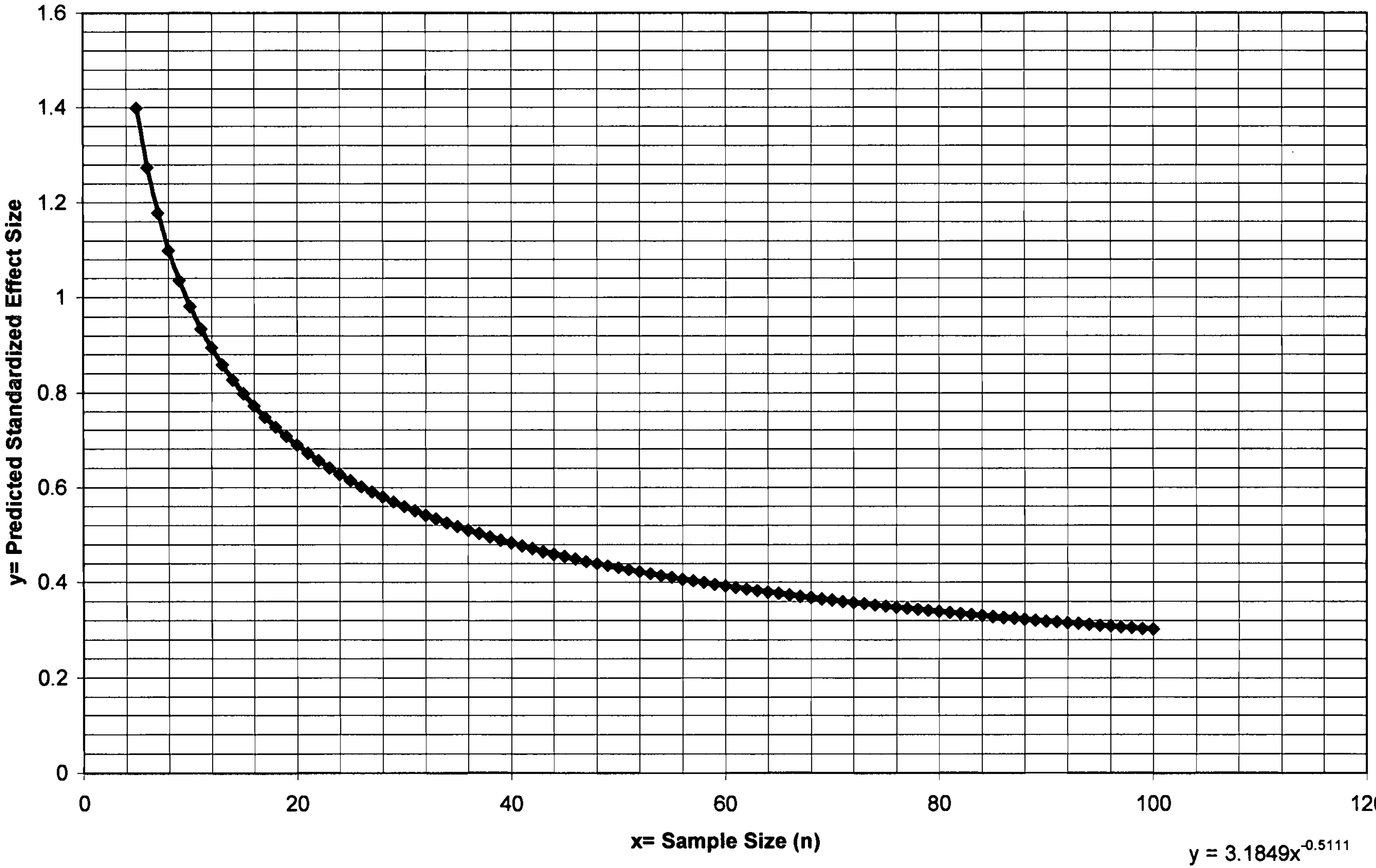


Figure 10. Predicted Standardized Effect Size by n for K= 10



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030764

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: EMPIRICALLY BASED CRITERIA FOR DETERMINING MEANINGFUL EFFECT SIZE	
Author(s): J. JACKSON BARNETTE + JAMES E. McLEAN	
Corporate Source:	Publication Date: 11/19/99

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: [signature]	Printed Name/Position/Title: J. JACKSON BARNETTE, ASSOC. PROF.	
Organization/Address: UNIVERSITY OF IOWA, 2511 STEINDLER IOWA CITY, IA 52242	Telephone: 319 335 8905	FAX: 319 335 6635
	E-Mail Address: JACK-BARNETTE@UIOWA.EDU	Date: 4/14/00

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com