DOCUMENT RESUME

ED 442 814 TM 031 234

AUTHOR Barnette, J. Jackson; McLean, James E.
TITLE Use of the Significance Test as a Protection against Spuriously High Standardized Effect Sizes: Introduction of the Protected Effect Size.
PUB DATE 2000-04-26
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Effect Size; Monte Carlo Methods; Simulation; *Statistical Significance; Test Use

ABSTRACT

The level of standardized effect sizes obtained by chance and the use of significance tests to guard against spuriously high standardized effect sizes were studied. The concept of the "protected effect size" is also introduced. Monte Carlo methods were used to generate data for the study using random normal deviates as the basis for sample means to be compared using one-way analysis of variance. Standardized effect sizes were generated for 5000 replications within each combination of number of groups from 2 to 10 and sample sizes from 5 to 100 in steps of 5, resulting in 900,000 total replications. Results indicate that the significance test does provide some protection against judging spuriously high standardized effect size values as being meaningful. Applying a statistical significance test does substantially reduce the proportion of standardized effect sizes that achieve small, medium, or large criteria levels by chance. Perhaps the arguments should not be about using either statistical significance testing or effect sizes. Instead, the focus should be on finding ways to combine the philosophies and methods of both to make decisions about group differences. (Contains 5 figures. 6 tables, and 36 references.) (SLD)

Use of the Significance Test as Protection Against Spuriously High Standardized Effect Sizes: Introduction of the Protected Effect Size

J. Jackson Barnette
University of Iowa

and

James E. McLean
University of Alabama at Birmingham

A paper presented at

The Annual Meeting of the American Educational Research Association

April 26, 2000

New Orleans LA

For further information, contact:

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

Rev. 5/3/00

Use of the Significance Test as Protection Against Spuriously High Standardized Effect Sizes: Introduction of the Protected 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 (1969, 1988) recommended specific levels of effect size for "small," "medium," and "large" effects. However, even Cohen acknowledged that these values are relative to the specific content and method in a given research situation. The purpose of this study is to determine the level of standardized effect sizes obtained by chance and the use of significance tests to guard against spuriously high standardized effect sizes. It introduces the concept of the "protected effect size."

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). 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 suggested general guidelines for levels of effect size. These are .2 for small effect, .5 for medium effect, and .8 for large effect.

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 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, 1993). 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 warned us about (1977, 1988, 1990).

Methodology

Monte Carlo methods were used to generate the data for this research using random normal deviates as the basis for sample means to be compared using one-way ANOVA. 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, resulting in 900,000 total replications. The standardized effect size was computed as the range of means divided by the root mean square error. In addition, the probability of the observed F statistic was evaluated using alpha values of .25, .10, .05, .01, and .001 so each observed effect size could be identified as being statistically significant in addition to achieving the general effect size criteria of .2 for a small effect size, .5 for a medium effect size, and .8 for a large effect size. Data were generated using a program written in double-precision Quick-BASIC. Analysis of the raw data was conducted using several routines of SAS®.

Results

General Properties of the Distribution of Standardized Effect Sizes

Table 1 presents the descriptive statistics for the distribution of standardized effect sizes for the total set of replications across the range of sample sizes and number of samples. Overall, the mean standardized effect size was 0.40639 (SD= 0.29182) with a range of 0 to 5.05513. The median standardized effect size was 0.34606. The distribution, as displayed in Figure 1, is clearly positively skewed (2.19932) and leptokurtic (7.76283).

As indicated in the total row of Table 2, more than 80% of the standardized effect sizes equaled or exceeded .2, the criterion for a small effect; 24.6% equaled or exceeded .5, the criterion for a medium effect size; 8.3% equaled or exceeded .8, the criterion for a large effect size; and 4.6% equaled or exceeded an effect size of 1.0. Clearly, it is not at all unusual to get a standardized effect size meeting the "small" effect size criterion, about one out of four meeting the "medium" effect size criterion, and almost one out of ten meeting the "large" effect size criterion.

Table 2 presents the proportion of those standardized effect sizes achieving selected criterion levels by number of samples. It is clear that as the number of samples increases, the proportion of standardized effect sizes achieving these cutoffs increases. Table 3 presents similar results but for differences in sample size. As sample size increases, the proportion of standardized effect size achieving the cutoffs decreases. Based on these trends, it could be predicted that the largest standardized effect sizes will be found in the condition of larger numbers of smaller sample sizes and the lowest standardized effect sizes will be found in the condition of a smaller number of larger sample sizes.

This prediction is verified in Table 4. The highest standardized effect sizes are found in the number of groups equals ten and sample size is 5 where the mean standardized effect size was 1.398 with a range of .427 to 3.184. The lowest standard effect sizes were found in the two-sample (t test) with large sample sizes where the mean standardized effect size was 0.112 with a range of 0 to 0.500. It is interesting to note that in the two-sample situation, with low sample size (10 or less) it is not at all unusual to get a standardized effect size well above the small criterion point.

If nothing else is clear at this point, it should be obvious that the current arbitrary and absolute criteria proposed by Cohen are far from being sensitive to the variation in standardized effect sizes as functions of the number and size of samples. Barnette and McLean (1999) examined this issue and provided preliminary equations which could be used to predict the mean standardized effect size that would occur by chance as functions of number of samples and sample size. They concluded that comparisons of observed effect sizes could be made using empirically determined standards based on characteristics such as number of samples and sample sizes rather than using fixed, arbitrary standards.

Use of the Significance Test as Protection

The primary question for this research is: Does the use of a significance test provide a way of reducing determination of spuriously high SES values as being meaningful? This issue will be addressed in two ways. First, an examination of proportions of effect sizes achieving or not achieving the three commonly used effect size criteria and being associated or not with statistical significance for the total of number and sample size configuration will be presented. Second, a more extensive analysis including number of sample and sample size combinations will be presented in the situation where alpha is set at .05 and the medium effect size criterion is used.

Table 5 presents the proportions of standardized effect sizes which do or do not achieve statistical significance at alpha levels of .5, .25, .10, .05, .01, and .001 and do or do not achieve/exceed the criteria for small, medium, and large effect sizes. In addition, results in Table 5 provide an indication of the reduction in effect sizes that could be considered to be meaningful when a significance test is applied as an additional criterion for judging effect size. In the following discussion, the term "protected effect size" will represent standardized effect sizes that meet the dual criteria of achieving the effect size criterion and statistical significance.

In the situation where the small effect size criterion is used, the proportion of standardized effect sizes achieving or exceeding .2 was .8037. Even the use of a significance test with alpha of .5 results in the proportion of 40.8% fewer than what might be determined as meaningful when no significance test is applied. The protected effect size proportion is .4758. When a significance test using an alpha of .25 is applied, there is a lower proportion by 69.1% and the protected effect size proportion is .2483. In these two cases, there were standardized effect sizes that were significant but did not achieve the small effect size criterion. However, this was a small proportion (.0235 when alpha was .5 and .0021 when alpha was .25). In every case, for the lower alpha levels, there were no incidences of standardized effect sizes that were statistically significant, but did not achieve the small effect size criterion. For the alpha levels less than .25, the protected effect size proportion was equal to alpha.

When alpha was set at .10, application of this as a criterion reduced the meaningful effect sizes by 87.6% with a protected effect size proportion of .1000. When alpha was set at .05, .1962 of the effect sizes were not statistically significant and did not achieve criterion level,

.7536 of the effect sizes were not significant but achieved the small effect size criterion, and all of those significant at .05 also achieved the criterion level with a proportion of .0501. These proportions are displayed in Figure 2. In reviewing Figure 2, it is important to remember that the standardized effect size is positively skewed and the two-dimensional view does not reflect that fact. In the .05 situation there was a reduction of 93.8% when the significance test was applied in addition to the small effect size criterion. When alpha was set at .01, there was a reduction of 98.8%; and when alpha was set at .001, there was a reduction of 99.9%. Thus, when looking for small effect sizes it is very easy to find them by chance in the ANOVA situation, but the use of at least a significance test with alpha set at .1 or lower removes about 88% or more of these effect sizes as potentially meaningful differences using a dual criteria decision model.

In the situation where the medium effect size criterion is used, the proportion of standardized effect sizes achieving or exceeding .5 was .2457. Even the use of a significance test with alpha of .5 results in the proportion of 24.3% fewer than what might be determined as meaningful when no significance test is applied. The protected effect size proportion is .1861. When using a significance test with an alpha of .25, there is a lower proportion by 50.9% and the protected effect size proportion is .1206. When alpha was set at .10, application of this as a criterion reduced the meaningful effect sizes by 75%.

When alpha was set at .05, .7395 of the effect sizes were not statistically significant and did not achieve criterion level, .2104 of the effect sizes were not significant but achieved the medium effect size criterion, .0148 were significant but did not achieve the medium effect size criterion, and .0353 were protected effect sizes. These proportions are displayed in Figure 3. In reviewing Figure 3, it is important to remember that the standardized effect size is positively skewed and the two-dimensional view does not reflect that fact. In the .05 situation there was a reduction of 85.6% when the significance test was applied in addition to the small effect size criterion.

When alpha was set at .01, there was a reduction of 96.4% with a protected effect size proportion of .0088; and when alpha was set at .001, there was a reduction of 99.6% with a protected effect size proportion of .0010, the value of alpha. Thus, in the alpha of .001 situation, the proportion of effect sizes that were significant at .001 and did not also achieve the medium effect size criterion was zero. When looking for medium effect sizes, it is relatively easy to find them by chance (about .25) in the ANOVA situation, but the use of a significance test with alpha set at .1 or lower removes 75% or more of these effect sizes as potentially meaningful differences using a dual criteria decision model.

In the situation where the large effect size criterion is used, the proportion of standardized effect sizes achieving or exceeding .8 was .0834. Even the use of a significance test with alpha of .5 results in the proportion of 20.4% fewer than what might be determined as meaningful when no significance test is applied. The protected effect size proportion is .0664. When using a significance test with an alpha of .25, there is a lower proportion by 46.8% and the protected effect size proportion is .0444. When alpha was set at .10, application of this as a criterion reduced the meaningful effect sizes by 72.1%.

When alpha was set at .05, .8800 of the effect sizes were not statistically significant and did not achieve criterion level, .0699 of the effect sizes were not significant but achieved the large effect size criterion, .0366 were significant but did not achieve the large effect size criterion, and .0135 were protected effect sizes. These proportions are displayed in Figure 4. In reviewing Figure 4, it is important to remember that the standardized effect size is positively skewed and the two-dimensional view does not reflect that fact. In the .05 situation there was a reduction of 83.8% when the significance test was applied in addition to the small effect size criterion.

When alpha was set at .01, there was a reduction of 95.7% with a protected effect size proportion of .0036; and when alpha was set at .001, there was a reduction of 99.4% with a protected effect size proportion of .0005. Thus, when looking for large effect sizes, about 9% may be found by chance in the ANOVA situation, but the use of at least a significance test with alpha set at .1 or lower removes 72% or more of these effect sizes as potentially meaningful differences using a dual criteria decision model.

Based on this analysis, it is clear that the use of the dual criteria model may rule out many standardized effect sizes as being meaningful when they could easily be attributed to chance. Even when selecting a very liberal value of alpha such as .25 or greater, there is evidence of a reduction of spuriously high standardized effect sizes. However, it is recommended that an alpha of .10 or lower be used in conjunction with attainment of an effect size criterion.

Since it is clear the standardized effect which may be observed by chance is a function of number of samples and sample size, the final part of this analysis looks at one of the situations, the medium standardized effect size with alpha set at .05, relative to the need for using a dual criteria decision model. Table 6 presents the proportions not achieving or achieving the medium effect size criterion and not significant or significant by number of samples and sample size. Examination of this table provides evidence that, as could be expected from the earlier results, the high number of small samples is the combination most in need of such a dual criteria. The small number of small samples situation is also in need of such a dual criteria model, but less so. Of all of the combinations, those in least need of such an adjustment are those with large sample sizes. In these, the number of samples is not as much a factor as when small sample sizes are used. Figure 5 presents the distributions of effect sizes for the four smallest-largest numbers of samples with the smallest-largest sample sizes. It could be assumed that similar findings would result from looking at the other combinations of effect size criteria and alpha levels.

Conclusion and Implication

The significance test does provide some protection against judging spuriously high standardized effect size values as being meaningful. Applying a statistical significance test does substantially reduce the proportion of standardized effect sizes that achieve small, medium, or large criteria levels by chance. This is clearly needed in cases where large numbers of small samples are used in experimental situations. A level of significance of .1 or lower does provide some minimal risk of committing a Type I error while reducing the number of standardized effect sizes that would be higher than criterion levels by chance.

Perhaps the arguments should not be in favor of using either statistical significance testing or effect sizes, but in finding ways, such as this, to combine the philosophies and methods of both to make decisions about meaningful group differences. It is certainly conceivable that a new probability could be determined that would use the combination of effect size and statistical significance under given conditions such as the number of samples and sample sizes. It should be possible to determine the probability of getting a given effect size and significance level, or a protected effect size probability, when the number of samples and sample sizes are known.

References

Barnette, J. J. & McLean, J. E. (1999, November). Empirically based criteria for determining meaningful effect size. A paper presented at the 1999 Annual Meeting of the Mid-South Educational Research Association, Point Clear, AL.

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.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*(5), 746-759.

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

Levin, J. R. (1993). Statistical significance testing from three perspectives. *Journal of Experimental Education, 61*(4), 378-381.

Levin, J. R. (1998). What if there were no more bickering about statistical significance tests? *Research in the Schools, 5*(2), 43-53.

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.

Robinson, D. L., & Levin, J. R. (1997). Reflections on statistical and substantive significance, with a slice of replication. *Educational Researcher, 26*(5), 21-26.

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.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods, 1*(2), 115-129.

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. Summary Statistic for Standardized Effect Size for Number of Samples from 2 to 10 and Sample Sizes of 5 to 100 in Steps of 5

n	900,000
Mean	0.40639
Median	0.34060
Mode	0.26625
Standard Deviation	0.29182
Minimum	0.00000
Maximum	5.05513
Skewness	2.19932
Kurtosis	7.76283
Standard Error of Mean	0.00031

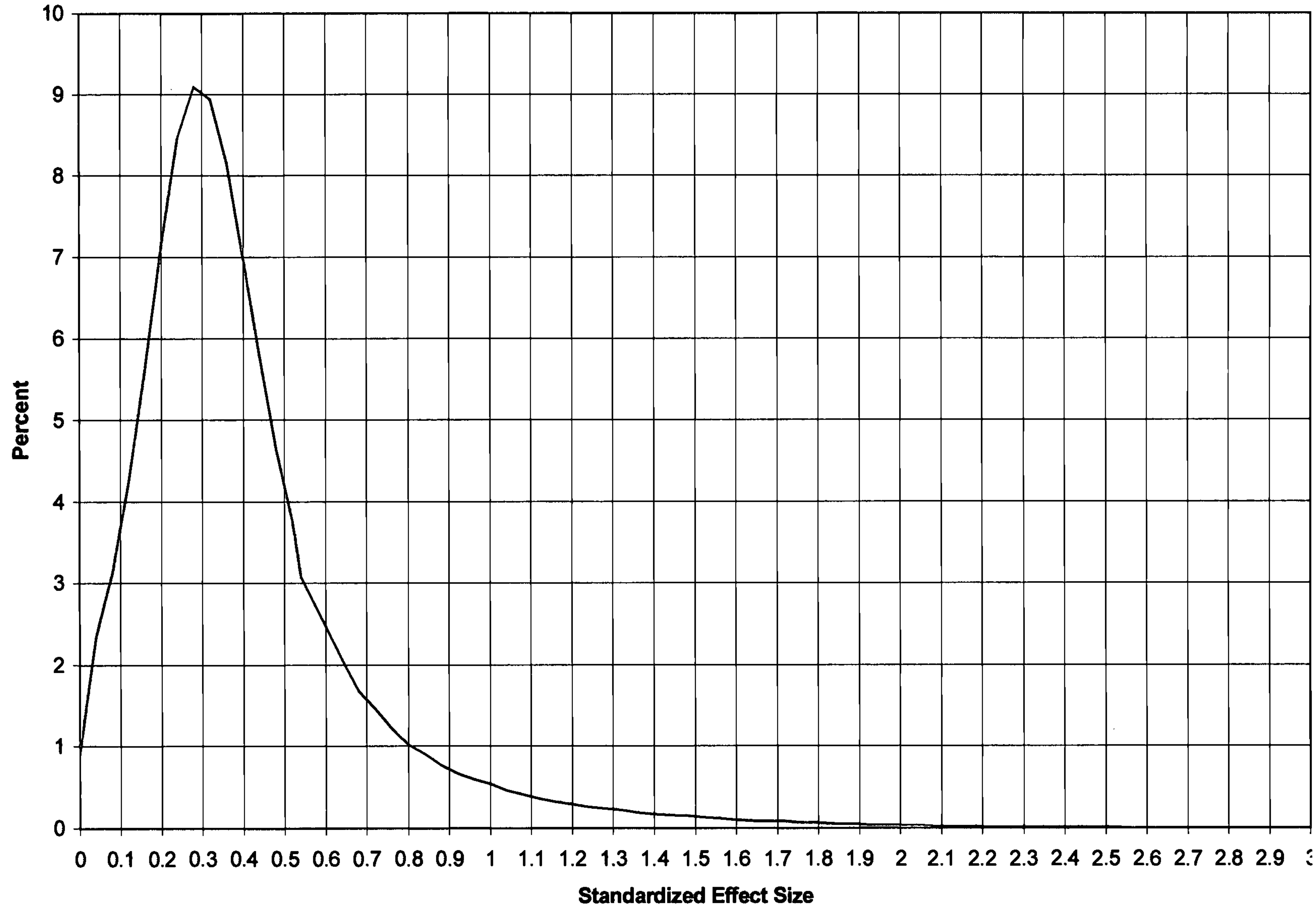

Figure 1. Frequency Polygon of Random Standardized Effect Sizes for k= 2-10 and n= 5-100

Table 2. Proportion of Standardized Effect Sizes Equal to or Above Selected Criterion Levels by Number of Samples, 100000 Replications per Number of Samples

Number of Samples, K	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
2	.6309	.3553	.1938	.1101	.0671	.0430	.0284	.0199	.0142	.0099	.0074	.0057	.0042	.0032	.0024
3	.8683	.5905	.3506	.2037	.1264	.0823	.0564	.0398	.0285	.0209	.0152	.0109	.0079	.0058	.0043
4	.9529	.7394	.4751	.2881	.1810	.1191	.0811	.0573	.0413	.0303	.0227	.0168	.0128	.0092	.0071
5	.9833	.8297	.5685	.3537	.2224	.1478	.1025	.0736	.0531	.0391	.0291	.0220	.0163	.0123	.0092
6	.9937	.8898	.6450	.4111	.2628	.1749	.1220	.0874	.0651	.0481	.0359	.0274	.0205	.0150	.0109
7	.9976	.9284	.7061	.4588	.2949	.1990	.1400	.1016	.0749	.0558	.0427	.0324	.0246	.0181	.0131
8	.9991	.9524	.7535	.5027	.3253	.2185	.1551	.1124	.0832	.0629	.0472	.0362	.0275	.0210	.0156
9	.9998	.9684	.7953	.5424	.3525	.2389	.1704	.1246	.0933	.0715	.0547	.0422	.0321	.0246	.0185
10	.9999	.9800	.8287	.5765	.3789	.2568	.1828	.1342	.1011	.0770	.0595	.0462	.0357	.0274	.0205
Total	.9362	.8038	.5907	.3830	.2457	.1645	.1154	.0834	.0616	.0462	.0349	.0267	.0202	.0152	.0113

Table 3. Proportion of Standardized Effect Sizes Equal to or Above Selected Criterion Levels by Sample Size, 45000 Replications per Sample Size

Sample Size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
5	.9856	.9661	.9444	.9155	.8786	.8334	.7794	.7168	.6461	.5670	.4859	.4076	.3326	.2652	.2056
10	.9774	.9462	.9032	.8468	.7732	.6790	.5658	.4490	.3371	.2375	.1590	.1003	.0603	.0343	.0184
15	.9708	.9299	.8691	.7823	.6655	.5297	.3835	.2535	.1506	.0829	.0408	.0206	.0092	.0040	.0016
20	.9662	.9136	.8327	.7142	.5602	.3904	.2397	.1296	.0608	.0253	.0103	.0040	.0012	.0005	.0001
25	.9600	.8934	.7918	.6410	.4590	.2787	.1452	.0636	.0234	.0076	.0020	.0005	.0001	.0000	.0000
30	.9549	.8793	.7574	.5754	.3712	.1946	.0856	.0298	.0091	.0024	.0006	.0001	.0000	.0000	.0000
35	.9491	.8635	.7160	.5077	.2915	.1313	.0487	.0146	.0038	.0008	.0001	.0000	.0000	.0000	.0000
40	.9446	.8457	.6766	.4426	.2260	.0874	.0265	.0060	.0012	.0001	.0000	.0000	.0000	.0000	.0000
45	.9406	.8331	.6407	.3900	.1748	.0584	.0152	.0027	.0003	.0000	.0000	.0000	.0000	.0000	.0000
50	.9363	.8114	.5967	.3334	.1342	.0387	.0079	.0014	.0002	.0000	.0000	.0000	.0000	.0000	.0000
55	.9323	.7949	.5609	.2872	.1014	.0242	.0048	.0007	.0001	.0000	.0000	.0000	.0000	.0000	.0000
60	.9276	.7800	.5233	.2451	.0767	.0160	.0028	.0003	.0000	.0000	.0000	.0000	.0000	.0000	.0000
65	.9252	.7638	.4853	.2072	.0573	.0100	.0013	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
70	.9183	.7448	.4506	.1753	.0427	.0065	.0007	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
75	.9152	.7276	.4215	.1484	.0319	.0041	.0003	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
80	.9118	.7121	.3895	.1250	.0239	.0030	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
85	.9075	.6922	.3552	.1040	.0176	.0021	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
90	.9039	.6758	.3283	.0848	.0123	.0011	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
95	.9002	.6605	.2977	.0742	.0102	.0008	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
100	.8956	.6412	.2738	.0602	.0064	.0003	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Total	.9362	.8038	.5907	.3830	.2457	.1645	.1154	.0834	.0616	.0462	.0349	.0267	.0202	.0152	.0113

Table 4. Standardized Effect Size Statistics by Number of Samples and Sample Size, 5000 Replications per Cell

n	Statistic	k= 2	k= 3	k= 4	k= 5	k= 6	k= 8	k= 10	Total*
5	M	0.5596	0.8171	0.9650	1.0814	1.1685	1.2949	1.3984	1.0987
	SD	0.4662	0.4699	0.4593	0.4475	0.4298	0.4111	0.3924	0.5068
	Min.	0.0001	0.0166	0.0488	0.1526	0.0976	0.2329	0.4274	0.0001
	Max.	3.5991	5.0551	3.5266	4.0714	3.3395	3.1073	3.1844	5.0551
10	M	0.3752	0.5475	0.6614	0.7445	0.8106	0.9041	0.9818	0.7601
	SD	0.2917	0.2979	0.2996	0.2901	0.2882	0.2766	0.2641	0.3416
	Min.	0.0001	0.0073	0.0453	0.0772	0.1303	0.1619	0.2240	0.0001
	Max.	1.9324	2.0711	2.1525	2.2750	2.1774	2.3311	2.1538	2.3311
15	M	0.2997	0.4454	0.5485	0.6070	0.6592	0.7372	0.7975	0.6193
	SD	0.2310	0.2399	0.2396	0.2337	0.2251	0.2172	0.2146	0.2752
	Min.	0.0001	0.0070	0.0320	0.0303	0.0822	0.1180	0.2088	0.0001
	Max.	1.6814	1.6528	1.6990	1.8542	1.5966	1.8071	1.7669	1.8542
20	M	0.2635	0.3868	0.4698	0.5220	0.5685	0.6382	0.6932	0.5346
	SD	0.2026	0.2060	0.2034	0.2000	0.1951	0.1882	0.1844	0.2358
	Min.	0.0002	0.0034	0.0249	0.0452	0.0623	0.1257	0.1828	0.0002
	Max.	1.1976	1.3665	1.4187	1.5536	1.4428	1.5091	1.5889	1.5889
25	M	0.2303	0.3365	0.4141	0.4674	0.5080	0.5738	0.6137	0.4764
	SD	0.1777	0.1804	0.1788	0.1761	0.1728	0.1636	0.1629	0.2110
	Min.	0.0000	0.0045	0.0337	0.0669	0.0877	0.1561	0.1541	0.0000
	Max.	1.1255	1.2940	1.2445	1.3300	1.3269	1.3141	1.3641	1.3641
30	M	0.2132	0.3112	0.3834	0.4281	0.4667	0.5202	0.5644	0.4358
	SD	0.1636	0.1699	0.1666	0.1639	0.1589	0.1521	0.1486	0.1926
	Min.	0.0000	0.0021	0.0233	0.0487	0.0793	0.1192	0.1336	0.0000
	Max.	1.1480	1.2115	1.1089	1.2140	1.1627	1.2639	1.2535	1.2639
40	M	0.1802	0.2666	0.3281	0.3696	0.4026	0.4483	0.4877	0.3755
	SD	0.1371	0.1416	0.1438	0.1397	0.1350	0.1298	0.1279	0.1657
	Min.	0.0000	0.0029	0.0213	0.0377	0.0259	0.1163	0.1110	0.0000
	Max.	0.9093	1.0157	0.9659	0.9683	0.9746	1.0279	1.0421	1.0421
60	M	0.1446	0.2203	0.2682	0.3017	0.3274	0.3684	0.3943	0.3066
	SD	0.1099	0.1171	0.1167	0.1137	0.1137	0.1070	0.1023	0.1355
	Min.	0.0000	0.0072	0.0068	0.0280	0.0284	0.1060	0.1105	0.0000
	Max.	0.7297	0.7194	0.9844	0.8214	0.8141	0.8684	0.8466	0.9844
80	M	0.1257	0.1901	0.2316	0.2608	0.2848	0.3181	0.3455	0.2658
	SD	0.0963	0.0996	0.1014	0.0970	0.0961	0.0917	0.0906	0.1173
	Min.	0.0000	0.0010	0.0085	0.0234	0.0347	0.0772	0.1030	0.0000
	Max.	0.6620	0.6924	0.7422	0.7442	0.6880	0.7648	0.7228	0.7648
100	M	0.1118	0.1685	0.2071	0.2328	0.2527	0.2840	0.3090	0.2371
	SD	0.0841	0.0881	0.0896	0.0869	0.0848	0.0823	0.0796	0.1044
	Min.	0.0000	0.0025	0.0104	0.0239	0.0352	0.0720	0.1032	0.0000
	Max.	0.5002	0.6262	0.6363	0.6667	0.7125	0.6244	0.6319	0.7125
Total*	M	0.1972	0.2924	0.3552	0.3987	0.4338	0.4852	0.5253	0.4064
	SD	0.2068	0.2381	0.2563	0.2691	0.2789	0.2943	0.3082	0.2918
	Min.	0.0000	0.0006	0.0044	0.0141	0.0161	0.0322	0.0779	0.0000
	Max.	3.5991	5.0551	3.5266	4.0714	3.3395	3.1073	3.1844	5.0551

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

Table 5. Proportions Not Achieving or Achieving Statistical Significance by Not Achieving or Achieving Effect Size Criterion

Alpha	Sign.	Effect Size Criterion	Effect Size		
			Small (.20) p(ach.)= .8038	Medium (.50) p(ach.)= .2457	Large (.80) p(ach.)= .0834
.50	$p \geq .50$	Not Achieved	.1727	.4410	.4837
		Achieved	.3279	.0596	.0170
	$p < .50$	Not Achieved	.0235	.3133	.4329
		Achieved	.4758	.1861	.0664
	Percent reduction using sign. test as additional criterion		40.8	24.3	20.4
.25	$p \geq .25$	Not Achieved	.1941	.6246	.7106
		Achieved	.5555	.1251	.0390
	$p < .25$	Not Achieved	.0021	.1297	.2060
		Achieved	.2483	.1206	.0444
	Percent reduction using sign. test as additional criterion		69.1	50.9	46.8
.10	$p \geq .10$	Not Achieved	.1962	.7157	.8398
		Achieved	.7037	.1842	.0601
	$p < .10$	Not Achieved	.0000	.0386	.0768
		Achieved	.1000	.0614	.0233
	Percent reduction using sign. test as additional criterion		87.6	75.0	72.1
.05	$p \geq .05$	Not Achieved	.1962	.7395	.8800
		Achieved	.7536	.2104	.0699
	$p < .05$	Not Achieved	.0000	.0148	.0366
		Achieved	.0501	.0353	.0135
	Percent reduction using sign. test as additional criterion		93.8	85.6	83.8
.01	$p \geq .01$	Not Achieved	.1962	.7530	.9101
		Achieved	.7937	.2369	.0798
	$p < .01$	Not Achieved	.0000	.0013	.0065
		Achieved	.0101	.0088	.0036
	Percent reduction using sign. test as additional criterion		98.8	96.4	95.7
.001	$p \geq .001$	Not Achieved	.1962	.7543	.9161
		Achieved	.8027	.2447	.0829
	$p < .001$	Not Achieved	.0000	.0000	.0005
		Achieved	.0010	.0010	.0005
	Percent reduction using sign. test as additional criterion		99.9	99.6	99.4

Table 6. Proportion of Mean Standardized Effect Sizes Not Achieving or Achieving or Exceeding "Medium Effect Size Criterion" and Not Significant or Significant at Alpha by Number of Samples and Sample Size, 5000 Replications per Cell

n	Medium Effect Size (.50)	Sign.	k= 2	k= 3	k= 4	k= 5	k= 6	k= 8	k= 10	Total*
5	Not Achieved	$p \geq .05$	.5454	.2718	.1418	.0720	.0316	.0092	.0014	.1214
		$p < .05$	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	Achieved or Exceeded	$p \geq .05$	.4038	.6798	.8082	.8786	.9220	.9434	.9522	.8298
		$p < .05$	.0508	.0484	.0500	.0494	.0464	.0474	.0464	.0489
10	Not Achieved	$p \geq .05$	.7128	.4942	.3288	.2070	.1320	.0504	.0172	.2277
		$p < .05$	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	Achieved or Exceeded	$p \geq .05$	.2370	.4562	.6188	.7450	.8166	.9022	.9376	.7228
		$p < .05$	.0502	.0496	.0524	.0480	.0514	.0474	.0452	.0495
15	Not Achieved	$p \geq .05$	.8166	.6380	.4602	.3506	.2598	.1356	.0700	.3345
		$p < .05$	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	Achieved or Exceeded	$p \geq .05$	.1360	.3118	.4840	.6010	.6932	.8192	.8786	.6154
		$p < .05$	.0474	.0502	.0558	.0484	.0470	.0452	.0514	.0502
20	Not Achieved	$p \geq .05$	.8690	.7322	.5944	.4910	.3892	.2376	.1480	.4398
		$p < .05$	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	Achieved or Exceeded	$p \geq .05$	.0724	.2198	.3568	.4596	.5618	.7158	.7980	.5101
		$p < .05$	.0586	.0480	.0488	.0494	.0490	.0466	.0540	.0500
25	Not Achieved	$p \geq .05$	.9170	.8142	.7006	.6018	.5094	.3490	.2540	.5410
		$p < .05$	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	Achieved or Exceeded	$p \geq .05$	.0332	.1414	.2554	.3458	.4420	.6028	.6974	.4107
		$p < .05$	.0498	.0444	.0440	.0524	.0486	.0482	.0486	.0483
30	Not Achieved	$p \geq .05$	.9390	.8602	.7630	.6968	.6076	.4698	.3592	.6288
		$p < .05$	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	Achieved or Exceeded	$p \geq .05$	.0090	.0822	.1820	.2498	.3432	.4844	.5926	.3199
		$p <.05$	.0514	.0578	.0550	.0534	.0492	.0458	.0482	.0513
40	Not Achieved	$p \geq .05$	.9502	.9328	.8770	.8300	.7730	.6722	.5632	.7715
		$p < .05$	.0210	.0008	.0008	.0000	.0000	.0000	.0000	.0025
	Achieved or Exceeded	$p \geq .05$	.0000	.0192	.0686	1218	.1748	.2808	.3848	.1782
		$p < .05$	.0288	.0472	.0536	.0482	.0522	.0470	.0520	.0478
60	Not Achieved	$p \geq .05$	.9514	.9472	.9470	.9390	.9144	.8772	.8474	.9092
		$p < .05$	.0418	.0334	.0174	.0104	.0064	.0060	.0032	.0140
	Achieved or Exceeded	$p \geq .05$	.0000	.0000	.0004	.0132	.0300	.0722	.1072	.0401
		$p < .05$	.0068	.0194	.0352	.0374	.0492	.0446	.0422	.0366
80	Not Achieved	$p \geq .05$	.9474	.9538	.9418	.9540	.9466	.9436	.9276	.9446
		$p < .05$	.0508	.0408	.0488	.0342	.0284	.0232	.0154	.0315
	Achieved or Exceeded	$p \geq .05$	.0000	.0000	.0000	.0000	.0004	.0052	.0222	.0048
		$p < .05$	.0018	.0054	.0094	.0118	.2460	.0280	.0348	.0191
100	Not Achieved	$p \geq .05$	.9530	.9532	.9476	.9514	.9508	.9452	.9412	.9492
		$p < .05$	.0468	.0462	.0490	.0454	.0442	.0458	.0432	.0444
	Achieved or Exceeded	$p \geq .05$	.0000	.0000	.0000	.0000	.0000	.0000	.0014	.0002
		$p < .05$	.0002	.0006	.0034	.0032	.0005	.0090	.0142	.0062
Total*	Not Achieved	$p \geq .05$	.9054	.8527	.8010	.7631	.7241	.6642	.6129	.7395
		$p < .05$	.0276	.0209	.0180	.0144	.0132	.0106	.0082	.0148
	Achieved or Exceeded	$p \geq .05$	.0446	.0980	.1472	.1875	.2256	.2863	.3372	.2104
		$p < .05$	.0225	.0284	.0338	.0349	.0372	.0389	.0418	.0353

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

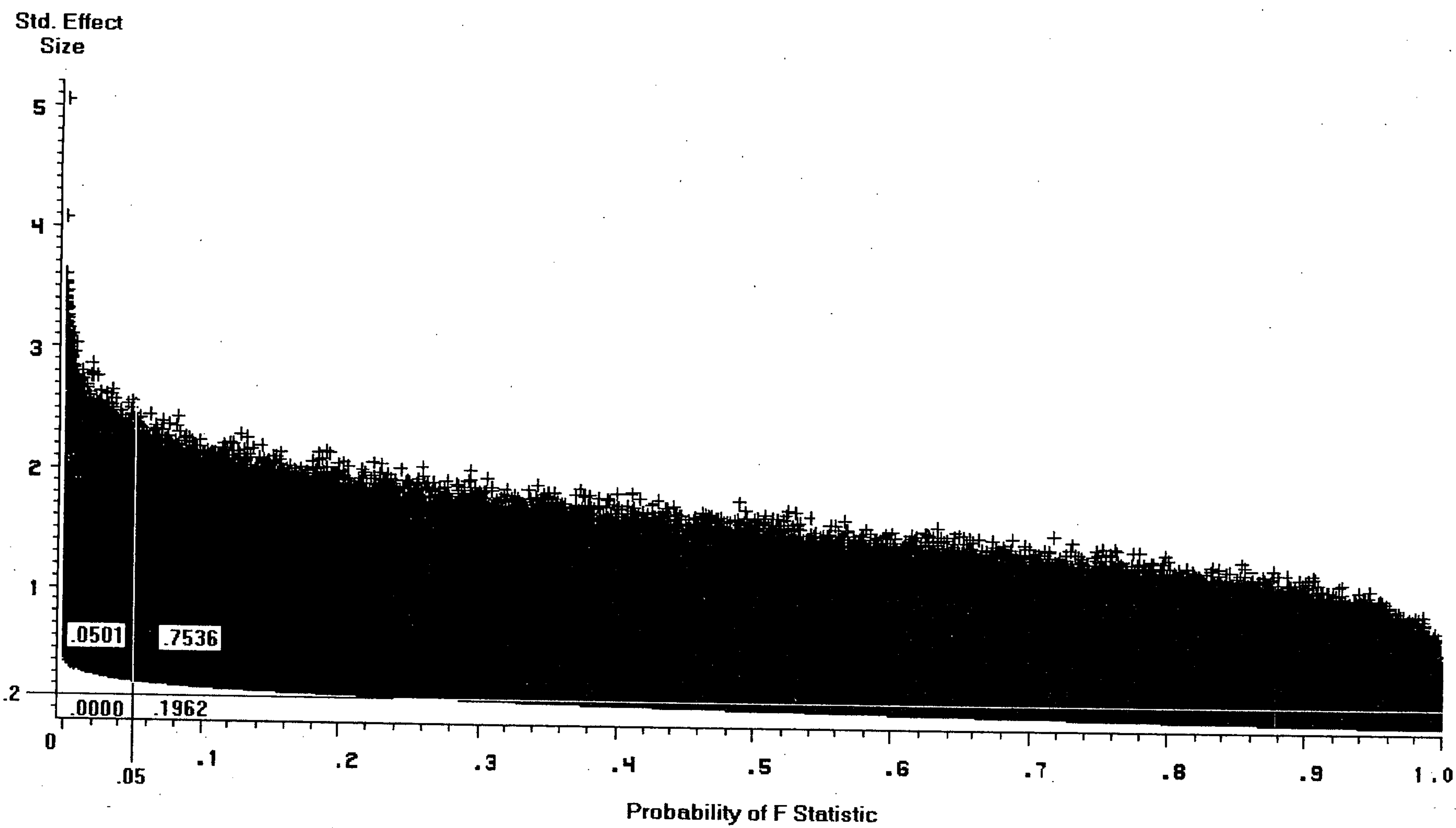

Figure 2. Proportion of Standardized Effect Sizes Achieving/Not Achieving Significance at .05 and Small Effect Size

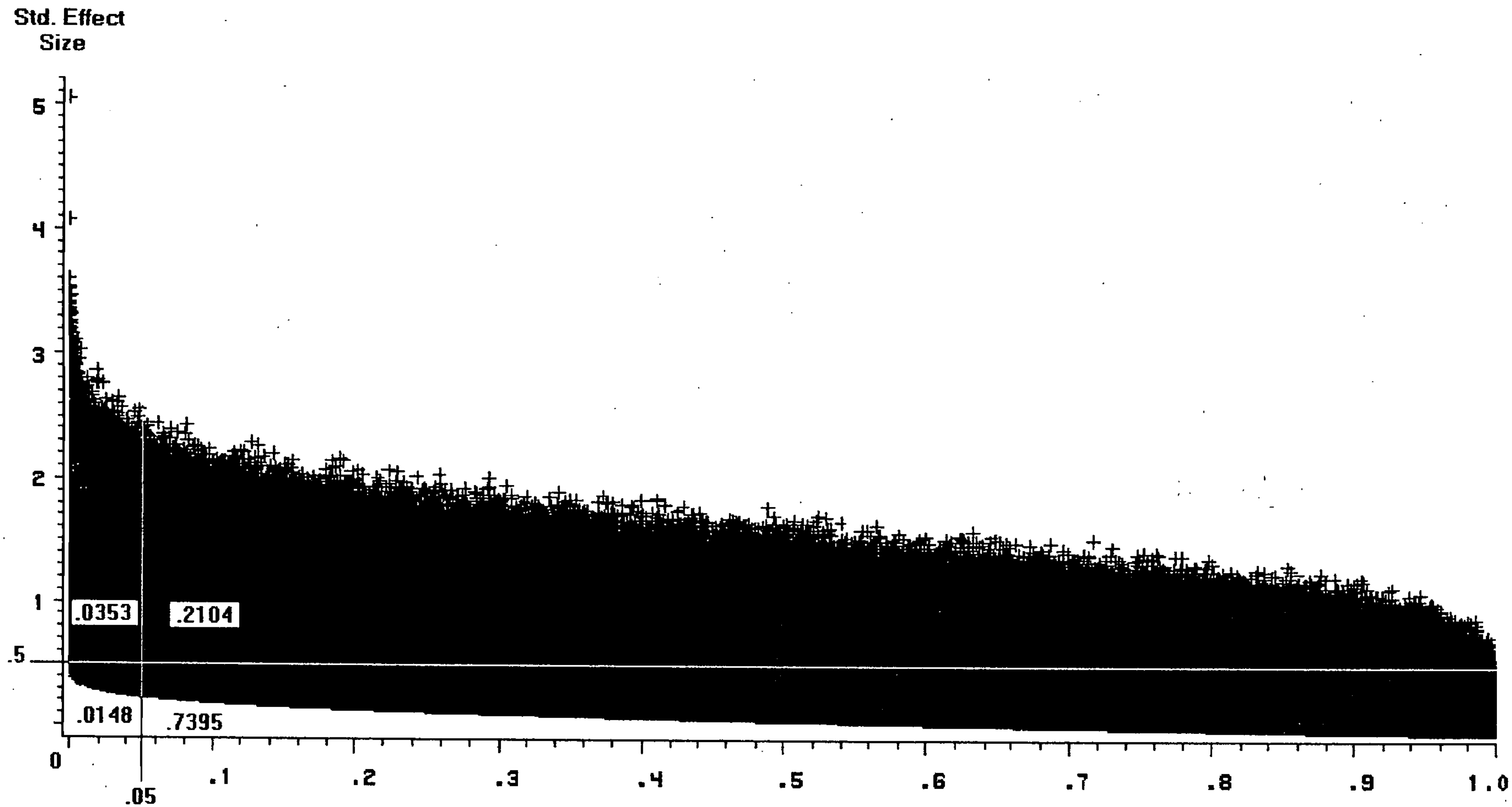

Figure 3. Proportion of Standardized Effect Sizes Achieving/Not Achieving Significance at .05 and Medium Effect Size

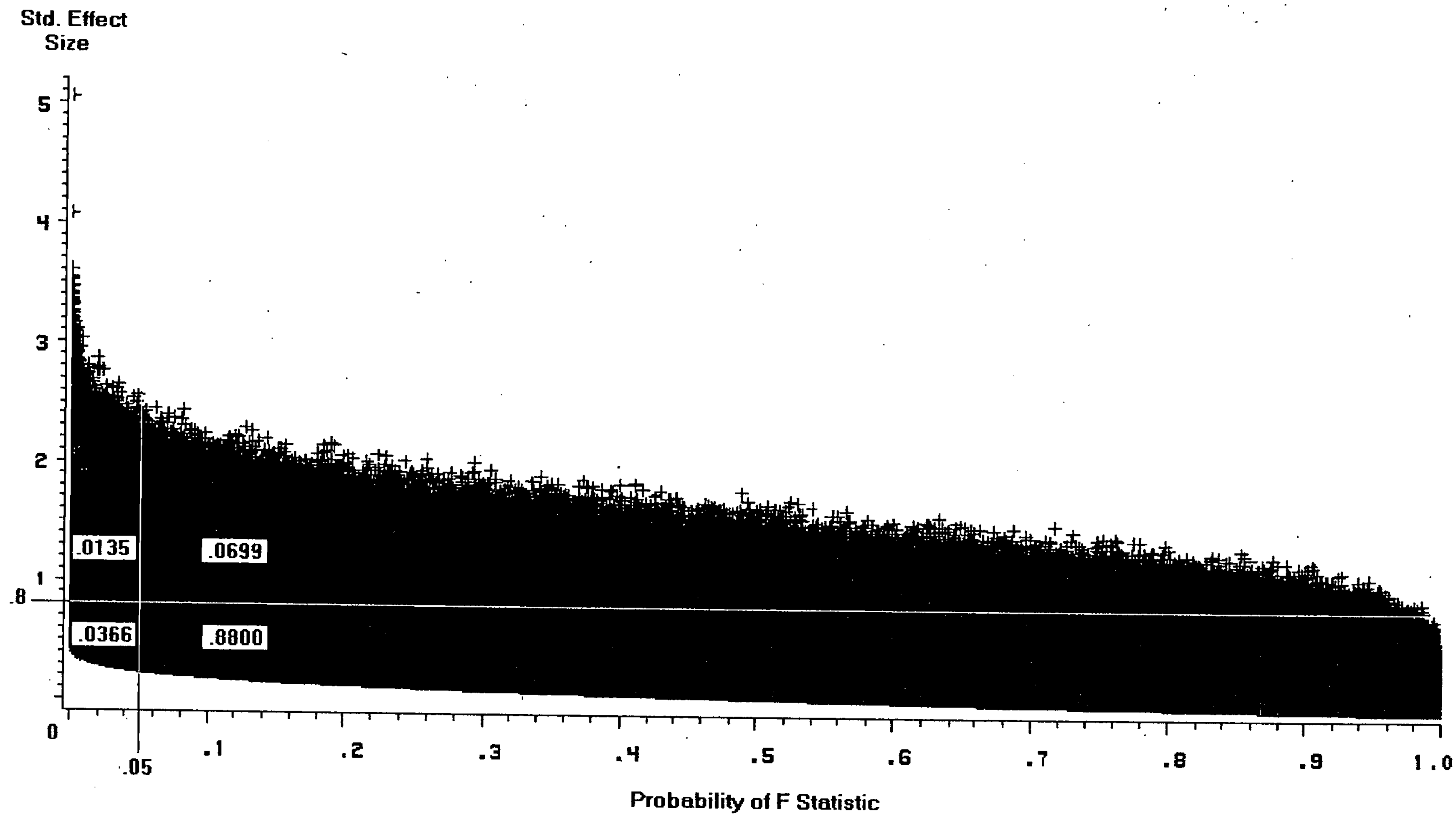

Figure 4. Proportion of Standardized Effect Sizes Achieving/Not Achieving Significance at .05 and Large Effect Size

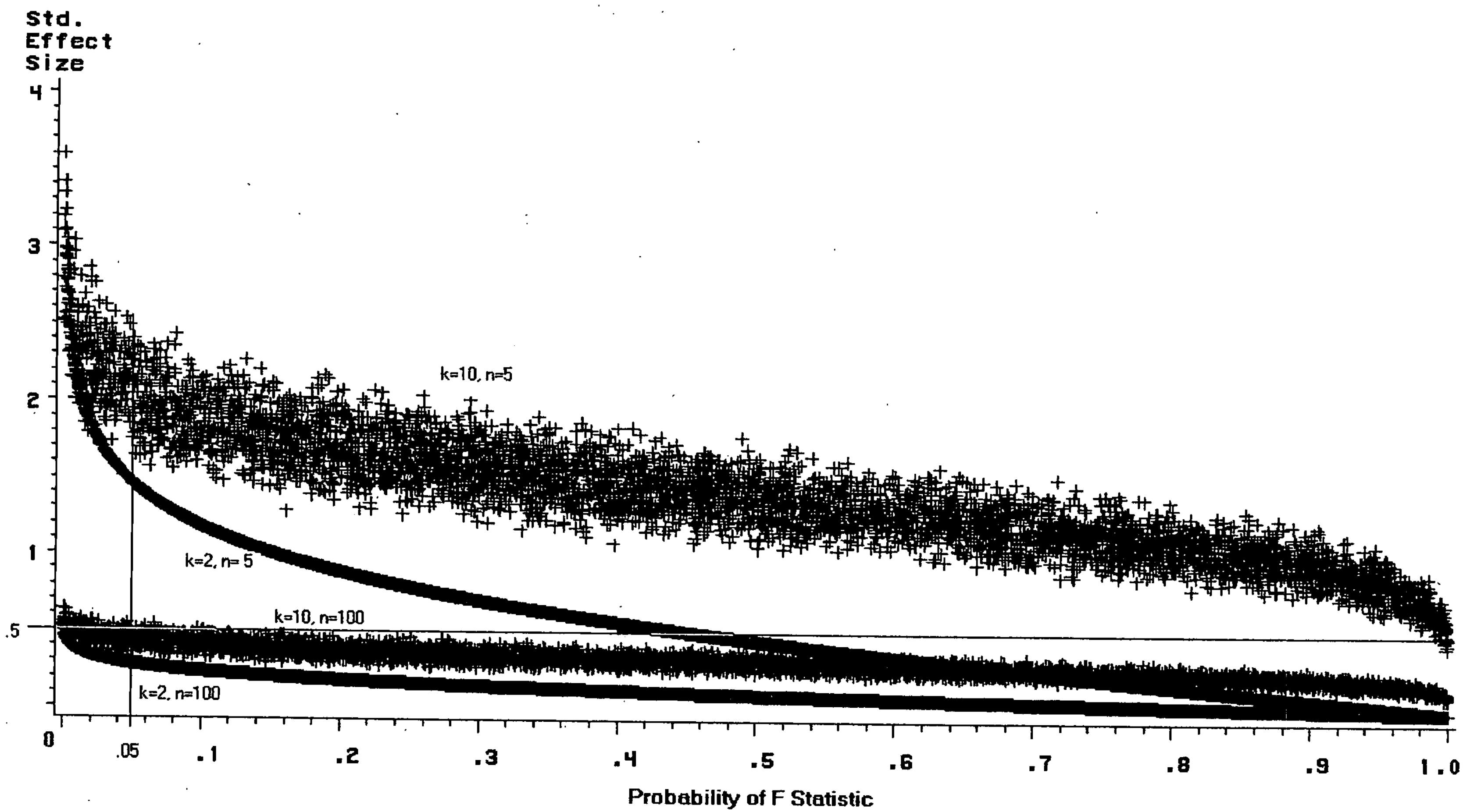

Figure 5. Distribution of Standardized Effect Sizes for Selected Number of Samples and Sample Size Configurations

ERIC®

REPRODUCTION RELEASE

TM031234

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: USE OF THE SIGNIFICANCE TEST AS PROTECTION AGAINST SPURIOUSLY HIGH STANDARDIZED EFFECT SIZES: INTRODUCTION OF THE PROTECTED EFFECT SIZE

Author(s): J JACKSON BARNETTE + JAMES E. MCLEAN

Corporate Source:

Publication Date: APRIL 2000

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)	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)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1	2A	
Level 1 ☑	Level 2 ☐	
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 relea... and dissemination in microfich... for ERIC archival collecti...	

Documents will be processed as indicated p...
If permission to reproduce is granted, but no box is ch...

THE UNIVERSITY OF IOWA

J. Jackson Barnette, Ph.D.
Associate Dean for Education and Student Affairs
Associate Professor
Community/Behavioral Health and Biostatistics
College of Public Health
2811B Steindler Bldg.
Iowa City, Iowa 52242-1008
319/335-8905 FAX 319/335-6635
jack-barnette@uiowa.edu

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: UNIV. OF IOWA 2811 STEINDLER BLDG. IOWA CITY IA 52242	Telephone: 319 335 8905	FAX: 319 335 6635
	E-Mail Address: JACK-BARNETTE @ UIOWA.EDU	Date: 5/3/00

(over)