Reporting and Interpreting Effect Size in Quantitative Agricultural Education Research

Joe W. Kotrlik, J. C. Atherton Alumni Professor
Louisiana State University
Heather A. Williams, Process Engineer
Folgers Coffee Company
M. Khata Jabor, Assistant Professor
Universiti Teknologi Malaysia

The Journal of Agricultural Education (JAE) requires authors to follow the guidelines stated in the Publication Manual of the American Psychological Association [APA] (2009) in preparing research manuscripts, and to utilize accepted research and statistical methods in conducting quantitative research studies. The APA recommends the reporting of effect sizes in quantitative research, when appropriate. JAE now requires the reporting of effect size when reporting statistical significance in quantitative manuscripts. The purposes of this manuscript are to describe the research foundation supporting the reporting of effect size in quantitative research and to provide examples of how to calculate effect size for some of the most common statistical analyses utilized in agricultural education research. Recommendations for appropriate effect size measures and interpretation are included. The assumptions and limitations inherent in the reporting of effect size in research are also incorporated.

Keywords: effect size, data interpretation, statistical significance

Introduction

"At present, too many research results in education are blatantly described as significant, when they are in fact trivially small and unimportant" (Carver, 1993, p. 287).

The term 'Effect Size' describes indices that measure the magnitude of treatment effects. 'Effect Size' differs from significance tests because it focuses on the meaning of the results and enables comparison between or among studies which further enables researchers to judge the practical significance of quantitative research results. "For the reader to appreciate the magnitude or importance of a study's findings, it is almost always necessary to include some measure of effect size in the Results section" (APA, 2009, p. 34). Additionally, effect size encourages a meta–analysis perspective thereby leading to the ability to compare between studies and demonstrate

repeatability of studies. Starting in January, 2010, the *Journal of Agricultural Education* (*JAE*) requires that "Authors MUST report effect sizes when reporting statistical significance for quantitative data analyses" (American Association for Agricultural Education [AAAE], 2010, p. 1).

Effect size has not been consistently reported in *Journal of Agricultural Education (JAE)* manuscripts in the 12 issues published in the last three years (2007–2009). The correct reporting of effect size has actually declined from a similar period 10 years ago (1997–1999) (Table 1). Out of the 119 manuscripts published in the last three years (2007–2009), effect size should have been reported in 55 manuscripts. An analysis of these 119 manuscripts revealed that effect size was reported correctly in 17 (30.9%) manuscripts and either not reported at all or reported incorrectly or inappropriately in the remaining 38 (69.1%) manuscripts. The data in Table 1 show that the correct reporting of effect size has declined from 36.8% in 1997–

1999 to 30.9% in 2006–2009. This comparison addressed *JAE*; however, it is probable that similar results would be obtained if this analysis was conducted for journals throughout the social sciences.

Table 1
Comparison of Effect Size Reporting and Interpretation in the Journal of Agricultural Education between 1997–1999 and 2007–2009

Year range	Total articles published (*N*)	Number of articles for which effect size should have been reported (*n*)	Effect size correctly reported and interpreted (*n*/%[a])	Effect size not reported, or incorrectly reported or interpreted (*n*/%[a])
1997–1999	87	38	14/36.8%	24/63.2%
2007–2009	119	55	17/30.9%	38/69.1%

[a]The *n* and % reported is based on the number of articles for which effect size should have been reported, as shown in column 3.

As discussed below, the reporting of effect size is an issue that has been strongly recommended by numerous researchers and journals. This article will focus on suggestions for reporting and interpreting effect size for inferential statistics commonly reported in manuscripts published in *JAE*. This manuscript is an updated and refocused version of a previously published manuscript (Kotrlik & Williams, 2003).

Theoretical Base

The concept of 'effect size' was first introduced as early as 1901 (Pearson). Interest in reporting 'effect size' has risen substantially in the last few decades and has become even more widespread in the research literature in recent years. Kirk (1996) cited the need to report effect size and an APA Task Force indicated that researchers should "Always provide some effect–size estimate when reporting a *p* value. . . . reporting and interpreting effect sizes in the context of previously reported effects is essential to good research" (Wilkinson & APA Task Force on Statistical Inference, 1999, p. 599).

Some common definitions for effect size include:

- ". . . a measure of the degree of difference or association deemed large enough to be of 'practical significance". . . (Cohen, 1962);

- ". . . the degree to which the phenomenon is present in the population . . . (Cohen, 1988, p. 9);
- ". . . estimate of the degree to which the phenomenon being studied . . . exists in the population . . ." (Hair, Black, Babin, Anderson, & Tatham, 2006, p. 2); and
- ". . . magnitude or importance of a study's findings . . ." (American Psychological Association, 2009, p. 34).
- Maxwell and Delaney (1990) indicated that two categories of measures of effect size are commonly utilized in the literature: measures of effect size (according to group mean differences), and measures of strength of association (according to variance accounted for).

The Importance of Effect Size
In 1901, Karl Pearson stated that statistical significance must be supplemented because it provides the reader with only a partial explanation of the importance or significance of the results (Kirk, 1996). Subsequently, Fisher (1925) proposed that, when reporting research findings, researchers should present measures of the strength of association or correlation ratios. Since these early observations, many researchers have promoted the use of effect size to complement or even replace statistical significance testing results, allowing the reader to interpret the results presented as well as providing a method of comparison of results between or among studies (Cohen, 1965, 1990, 1994; Hair et al., 2006; Kirk, 1996, 2001;

Thompson, 1998, 2002). Effect size can also be valuable in characterizing the degree to which sample results diverge from the null hypothesis (Cohen, 1988, 1994). Therefore, reporting effect size allows a researcher to judge the magnitude of the differences between or among groups, which increases the researcher's capability to compare current research results to previous research and judge the practical significance of the results derived.

JAE requires authors to prepare their manuscripts in accordance with the *Publication Manual of the American Psychological Association* (2009) which provides guidance for authors regarding effect size (AAAE, 2010). The emphasis on effect size by the APA was preceded by an APA Task Force's earlier recommendation that strongly encouraged researchers to report effect sizes such as Cohen's *d*, Cohen's *f*, eta^2, or adjusted R^2 (Wilkinson & APA Task Force on Statistical Inference, 1999, p. 599).

Ample support for APA's recommendations regarding the reporting of effect size is available in the research literature. One example is Fan (2001) who indicated that good research presented both statistical significance testing results and effect sizes. Baugh (2002) stated that "Effect size reporting is increasingly recognized as a necessary and responsible practice" (p. 255). It is the researcher's duty to adhere to stringent analytical and reporting methods in order to ensure the proper interpretation and application of research results. The reporting of effect size is a part of this duty.

Why Report Effect Size in Addition to Statistical Significance?

Reporting effect size in addition to reporting statistical significance is important because many researchers assume a *p* value provides an indicator of both statistical and practical significance. Attention to the misuse of statistical testing spans the research literature from Cohen (1988) to Kline (2009) and Thompson (2009), and will likely continue for years to come. Misuse begins as early in the study design as selection of the *alpha* value and continues through the interpretation of the results of the selected statistical test.

Researchers set an *alpha* value (or probability of Type I error) based on the amount of risk one is willing to accept that one will

incorrectly reject the null hypothesis; as well as previous research in their field (Hair et al., 2006; Mendenhall, Beaver, & Beaver, 1999). Generally, Type II error is not considered, thereby risking the practical significance implications of one's study. Once the statistical test is conducted, Nickerson (2000), stated that many researchers misinterpret the results by believing that a small value of *p* means a treatment effect of large magnitude, and that statistical significance means theoretical or practical significance. A researcher must remember that "Statistical significance testing does not imply meaningfulness," (Olejnik & Algina, 2000, p. 241); and that statistical significance testing determines the probability of obtaining the sampling outcome by chance, while it is effect size that addresses practical significance or meaningfulness (Fan, 2001). Additionally, Kirk (2001) reminds researchers of the reliance of statistical significance on sample size, but notes that effect size assists one in the interpretation of results and thereby making trivial effects harder to ignore, and furthering the ability of a researcher to decide whether results are practically significant (Kirk, 2001).

These arguments against interpretation of a *p* value to denote practical significance lead researchers to recognize the value of including effect size measures in statistical testing execution, interpretation, and reporting. As an example of the application of these arguments, the reporting of effect size in a recent agricultural education study would have led the researchers to use effect size in addition to the results of a statistically significant *t*–test as the basis for drawing their conclusions. The authors compared students' perceptions of agriculture in schools by whether the school had an agriculture program. The data in Table 2 show that even though the *t*–test was statistically significant ($t = 2.00$, $p = .046$, $df = 1,767$), Cohen's effect size value ($d = .10$) did not meet Cohen's minimum standard ($d \geq .20$) to be called a "small" effect size. This additional information may have led the researchers to conclude that the differences have **negligible** practical significance and no substantial recommendations may be appropriate based on the results of this *t*–test.

As can be seen from the research literature regarding effect size and the example above, it is the researcher's responsibility to select the most appropriate sample size statistical test(s),

properly set the *alpha* value, properly select the appropriate effect size measure, determine the most appropriate interpretation method, clearly report all results, and base conclusions and recommendations on the overall results (i.e., the "big picture" based on **BOTH** the *p* value interpretation **AND** effect size interpretation). These actions increase the ability to determine

not only statistical significance but also practical significance, further adding to the ability of the researcher to determine whether the outcome may or may not have occurred by chance (Capraro 2004; Carver, 1993; Fagley & McKinney, 1983; Fan, 2001; Kline (2009); Robinson & Levin, 1997; Shaver, 1993; Thompson, 1996, 1998, 2009).

Table 2

Comparison of General Agriculture Perceptions by Students in Schools with an Agriculture Program versus Students in Schools with No Agriculture Program (N = 1,953)

	Agriculture program		No agriculture program					
	M	*SD*	*M*	*SD*	*t*	*df*	*P*	Cohen's *d*
Student perceptions of agriculture	20.11	2.68	19.86	2.55	2.00	1767	.046	.10

Note. The data in this table were taken from a recently published agricultural education research manuscript. Adapted with permission.

Cautions Applicable to Effect Size Interpretation

Just as a researcher must be cautious regarding the violation of assumptions when computing a parametric test statistic, one must recognize that effect size measures are also sensitive to violations of assumptions – specifically, non–normality and heterogeneity (Leech & Onwuegbuzie, 2002). With this in mind, researchers must first ensure that the assumptions of the statistical test are satisfied when conducting their analyses; then, carefully select the most appropriate effect size measure. The researcher should select effect size measures after testing for the violation of assumptions and before the actual execution of parametric or non–parametric tests. Those researchers using non–parametric test statistics should ensure the effect size measure selected is not a parametric effect size measure (Leech & Onwuegbuzie, 2002). Additionally, researchers should be cautious completing analyses with small sample sizes as these have the potential to influence the results of effect size calculation.

The next caution in reporting effect size is interpretation. In his initial publication that proposed an interpretation of effect size measures, Cohen (1988) did not anticipate such wide utilization and acceptance. However, Cohen's book and other publications such as Davis (1971) and Hinkle, Wiersma, and Jurs (1979), have permeated the education literature from both a referential and debate stance; and

provide points of reference to assist a researcher in deciding how to interpret the magnitude of the results of their study. For the purpose of this discussion, methods for determining effect size for the most commonly used statistical analyses are presented below in the "Measures of Effect Size" section of this manuscript. The authors note, however, that one should not rely solely on this manuscript (or any one publication) as each provides only a limited perspective of the appropriate use of effect sizes. Additionally, effect sizes presented are generally available in current statistical programs such as SAS and SPSS and on the Internet; other effect size measures are available for use but must be calculated by hand.

Effect Size Measures

Effect size measures have been divided into two families (*d* family and *r* family) by Rosenthal (1994). This division assists in understanding of appropriate application of effect size measures. The *d* family is most often associated with variations on standardized mean differences, while the *r* family is expressed in terms of the correlation coefficient (*r* or r^2). With some parametric test statistics, both may be used as appropriate effect size estimates. For the purposes of this manuscript, those effect size measures commonly applied to the parametric tests most frequently published in *JAE* will be

discussed. These span the *r* and *d* families of indices.

The previously mentioned analysis of recent *JAE* articles revealed the most commonly used statistical tests. These included multiple regression, independent *t*-test, ANOVA, bivariate correlation, ANCOVA, Chi–square, and factor analysis (Table 3). Other analyses used that are common to most social science research include paired *t*-test and point–biserial correlation. Each of these statistical tests has associated effect size measures that are most appropriate. Possible selections of effect size measures are presented in Table 3, and their use and interpretation is presented in the discussion below. Please note that although some effect size measures are the same for different tests (e.g., Cohen's *d*), the denominator is calculated differently as discussed following Table 3.

Table 3
Statistical Tests Reported in the Journal of Agricultural Education

Statistical test	Reported use in *JAE* 2007 – 2009	Potential effect size measures[a]
Multiple regression	7	Multiple regression coefficient (R^2)
Independent *t* test[b]	6	Cohen's *d*, Hedges's *g*, Glass's delta
ANOVA	5	Cohen's *f*, Omega squared
Bivariate correlation	4	Correlation coefficient (Pearson's *r*, Spearman's *rho*)
ANCOVA	3	Cohen's *f*, Omega squared
Chi square	2	Phi coefficient (2x2 table), Cramer's V (larger than 2x2 table)
Paired *t* test[b]	1	Cohen's *d*
Point biserial correlation	1	Correlation coefficient (point bi–serial)

[a]See Table 2 for potential effect size descriptors. [b]The formulas used to calculate Cohen's *d* differ for paired and inferential *t*–tests (Cohen, 1988).

Independent t–tests

Cohen's *d* statistic is a common measure to estimate effect size for independent samples *t*-tests (Cohen, 1988). If the statistical analysis program utilized by the researcher does not calculate Cohen's *d*, one will need the following formulas to calculate the pooled standard deviation and the Cohen's *d* statistic:

Pooled standard deviation = *sqrt* $[((n_1-1)s_1^2 + (n_2-1)s_2^2) / ((n_1-1)+((n_2-1))]$

Then, Cohen's *d* = Difference between sample means / Pooled standard deviation

For formula assistance, Larry Beckley provides a web–based effect size calculator (Cohen's *d*) for independent samples *t*–tests which may be found at http://www.uccs.edu/~faculty/lbecker/ index.html. Hedges's *g* and Glass's delta are also commonly referenced measures to estimate effect size for independent *t*–tests (Kline, 2009; Rosenthal & Rosnow, 1991). Hedge's *g* is most appropriate for very small samples.

Paired t–tests

Cohen's *d* is also applicable to estimate effect size for paired samples *t*–tests. Paired sample *t*–tests compare group means when the two groups are correlated in various research designs (e.g., matched pairs, repeated measures, before–after). The denominator should be calculated using the original standard deviations (Dunlop, Cortina, Vaslow, & Burke, 1996). Researchers should be cautious when using web–based effect size calculators to ensure that the calculator they select is appropriate for paired *t*–tests.

Analysis of Variance and Analysis of Covariance

Both Cohen's *f* (Cohen, 1988) and Omega squared (ω^2) are common methods of reporting effect sizes for analyses of variance when utilizing ANOVA and ANCOVA. Both provide an estimate of the proportion of variance explained by the categorical variable, with Cohen's *f* estimating the proportion of variance explained for the sample, while Omega squared estimates the proportion of variance explained for the population. To calculate Cohen's *f*, calculate *Eta* squared (η^2) first:

$$\eta^2 = SS_{Between}/SS_{Total}$$

Then, use the following formula to calculate Cohen's *f*:

$$\text{Square root of } (\eta^2/1 - \eta^2)$$

Omega squared (ω^2) can be calculated as follows:

$$\omega^2 = \frac{SS_{Between} - (k-1)MS_{Within}}{SS_{Total} + MS_{Within}}$$

In the formula for ω^2, k = number of groups. The sum of square and mean square information is provided by most statistical programs.

> *Note*: If assumptions of equal sample size and homogeneity of variance are violated, effect size will be overestimated (Volker, 2006); therefore, caution should be used in interpreting and reporting the effect size measure. Use the descriptors in Table 4 to interpret these coefficients.

Correlations

Perhaps the simplest measures and reporting of effect sizes exist for correlations. The correlation coefficient itself is a measure of effect size. The most commonly used statistics for parametric correlations are Pearson's *r* and Spearman's *rho* (r_s); and the most commonly used statistic for nonparametric correlations is the point biserial correlation (r_{pb}). The practical importance of correlation coefficients must be interpreted descriptors for correlation coefficients. Several sets of descriptors for correlation coefficients are presented in Table 4.

Non–parametric Measures

The *Phi* coefficient is commonly used to estimate the magnitude of association in 2 x 2 contingency tables. *Phi* is a Pearson product–moment coefficient calculated on two nominal–dichotomous variables for which the categories of both variables have been coded 0 and 1. Cramer's *V* is commonly used to describe the magnitude of association between categorical variables for a contingency table larger than 2 x 2. SPSS, SAS and other statistical analysis programs will calculate either the *Phi* or Cramer's *V* coefficients. Use the descriptors in Table 4 to interpret these coefficients.

Regression

An effect size measure for simple or multiple regression is the regression coefficient, R^2. Most statistical analysis programs calculate this coefficient which represents the proportion of the dependent variable's variance that is explained by the independent variable(s). The effect size of the calculated R^2 may be interpreted using the set of descriptors proposed by Cohen (1988, see Table 4).

Table 4

Descriptors for Reporting and Interpreting Effect Size in Quantitative Research

Reference	Effect size statistic	Values	Interpretation of effect size
Cohen, 1988		.10	Small effect size
		.30	Medium effect size
	Cramer's *Phi* or Cramer's *V* for nominal data	.50	Large effect size
Rea & Parker, 1992		.00 and under .10	Negligible association
		.10 and under .20	Weak association
		.20 and under .40	Moderate association
		.40 and under .60	Relatively strong association
		.60 and under .80	Strong association
		.80 and under 1.00	Very strong association
Cohen, 1988	Cohen's *d* for independent *t*-tests	.20	Small effect size
		.50	Medium effect size
		.80	Large effect size
	Cohen's *f* for ANOVA and ANCOVA	.10	Small effect size
		.25	Medium effect size
		.40	Large effect size
Cohen, 1988	R^2 for multiple regression	.0196	Small effect size
		.1300	Medium effect size
		.2600	Large effect size
Keppel, 1991	Omega squared (ω^2) for ANOVA, ANCOVA	.01	Small effect
		.06	Medium effect
		.15	Large effect
Kirk, 1996		.010	Small effect size
		.059	Medium effect size
		.138	Large effect size
Davis, 1971[a]		.70 or higher	Very strong association
		.50 to .69	Substantial association
		.30 to .49	Moderate association
		.10 to .29	Low association
		.01 to .09	Negligible association
Hinkle, Wiersma, & Jurs, 1979[ab]	Correlation coefficients	.90 to 1.00	Very high correlation
		.70 to .90	High correlation
		.50 to .70	Moderate correlation
		.30 to .50	Low correlation
		.00 to .30	Little if any correlation
Hopkins (1997)[a]		.90 to 1.00	Nearly, practically, or almost: perfect, distinct, infinite
		.70 to .90	Very large, very high, huge
		.50 to .70	Large, high, major
		.30 to .50	Moderate, medium
		.10 to .30	Small, low, minor
		.00 to .10	Trivial, very small, insubstantial, tiny, practically zero

Note. Table adapted from Kotrlik & Williams, 2003. Adapted with permission.
[a]Several authors have published guidelines for interpreting the magnitude of correlation coefficients.
[b]Note the more stringent nature of these descriptors when compared to Davis (1971) and Hopkins (1997).

Discussion

This manuscript does not identify all available effect size measures. The effect size measures selected for presentation address only the most commonly used statistical analyses in manuscripts published in the *JAE*. It is not difficult to calculate and report effect sizes and the references provided at the end of this article will serve as a good starting point for authors attempting to identify appropriate measures of effect size for other types of statistical analyses.

The guidelines referenced in this article for the interpretation of effect sizes should be taken as general guidelines to follow if previous findings and knowledge of the area studied do not exist. If previous findings or knowledge of the area studied exist, they should be used in consort with the statistical significance results and the calculated effect size to interpret the practical importance of the findings. Thompson (2000) supported these cautionary words by stating, ". . . it must be emphasized that if we mindlessly invoke Cohen's rule of thumb, contrary to his strong admonitions, in place of the equally mindless consultation of *p* value cutoffs such as .05 and .01, we are merely electing to be thoughtless in a new metric" (Thompson, 2000, ¶18)

The authors' main purpose for writing this article was their hope that *JAE* authors will improve the reporting of their research by including and interpreting effect size measures when appropriate. It is the hope of the authors that this article will serve as a useful resource for agricultural education researchers and that the reporting of effect sizes will strengthen the quantitative research articles published in *JAE*.

References

American Association for Agricultural Education. (2010). *Manuscript submission guidelines* [for the *Journal of Agricultural Education*]. Retrieved from http://www.jae-online.org/images/stories/Other/Manuscript_Preparation_Guidelines.pdf

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Baugh, F. (2002). Correcting effect sizes for score reliability: A reminder that measurement and substantive issues are linked inextricably. *Educational and Psychological Measurement, 62*(2), 254–263. doi: 10.1177/0013164402062002004

Capraro, R. M. (2004). Statistical significance, effect size reporting, and confidence intervals: Best reporting strategies. *Journal for Research in Mathematics Education*, *35*(2), 57–62.

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

Cohen, J. (1962). The statistical power of abnormal–social psychological research: A review. *Journal of Abnormal Psychology, 65*, 145–153.

Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology*. New York, NY: Academic Press.

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

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

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

Cohen, J. (1994). The earth is round (*p<.05*). *American Psychologist, 49,* 997–1003.

Davis, J. A. (1971). *Elementary survey analysis*. Englewood Cliffs, NJ: Prentice–Hall.

Dunlop, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta–analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177.

Fagley, N. S., & McKinney, I. J. (1983). Review bias for statistically significant results: A reexamination. *Journal of Counseling Psychology, 30,* 298–300.

Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. *The Journal of Educational Research, 94*(5), 275–282.

Fisher, R. A. (1925). *Statistical methods for research workers*. London: Oliver & Boyd.

Hair, J. F, Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). NJ: Prentice Hall.

Hinkle, D. E., Wiersma, W., & Jurs, S.G. (1979). *Applied statistics for the behavioral sciences*. Chicago, IL: Rand McNally College Publishing.

Hopkins, W. G. (1997). *New view of statistics*. Retrieved from http://www.sportsci.org/resource/stats/effectmag.html.

Keppel, G. (1991). Design and analysis: A researcher's handbook (3rd ed.) Englewood Cliffs, NJ: Prentice–Hall.

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

Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement, 61*(2), 213–218.

Kline, R. B. (2009). *Becoming a behavioral science researcher: A guide to producing research that matters.* New York, NY: Guilford Press.

Kotrlik, J. W., & Williams, H. A. (2003). The incorporation of effect size in information technology, learning, and performance research. *Information Technology, Learning, and Performance Journal, 21*(1), 1–7.

Leech, N. L., & Onwuegbuzie, A. J. (2002, November). *A call for greater use of nonparametric statistics*. Paper presented at the Annual Meeting of the Mid–South Educational Research Association, Chattanooga, TN.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (1999). *Introduction to probability & statistics*. Pacific Grove, CA: Brooks/Cole.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5,* 241–301.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25,* 241–286. doi: 10.1006/ceps.2000.1040

Pearson, K. (1901). On the correlation of characters not quantitatively measurable. *Philosophical Transactions of The Royal Society of London, 195,* 1–47.

Rea, L. M., & Parker, R. A. (1992). *Designing and conducting survey research*. San Francisco, CA: Jossey–Bass.

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

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook for research synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.

Rosenthal, R., & Rosnow, R. L. (1991). *Effects of behavioral research: Methods and data analysis* (2nd ed.). New York, NY: McGraw–Hill.

Shaver, J. P. (1993). What statistical significance testing is and what it is not. *The Journal of Experimental Education, 61,* 293–316.

Thompson, B. (1996). AERA editorial polices regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*(2), 26–30.

Thompson, B. (1998). In praise of brilliance: Where that praise really belongs. *American Psychologist, 53,* 799–800.

Thompson, B. (2000). *A suggested revision to the forthcoming 5th edition of the APA Publication Manual*. College Station, TX: Texas A&M University. Retrieved from http://www.coe.tamu.edu/~bthompson/apaeffec.htm

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 25–32. doi: 10.3102/0013189X031003025

Thompson, B. (2009). Computing and interpreting effect sizes in educational research. *Middle Grades Research Journal, 4*(2), 11–24

Volker, M. (2006). Reporting effect size estimates in school psychology research. *Psychology in the Schools, 43*(6), 653–672. doi: 10.1002/pits.20176

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604.

JOE W. KOTRLIK is the J. C. Atherton Alumni Professor in the School of Human Resource Education and Workforce Development, 142 Old Forestry Building, Louisiana State University, Baton Rouge, LA 70803–5477. Telephone: 225–578–5753, kotrlik@lsu.edu

HEATHER A. WILLIAMS directs an employee development project for Folgers Coffee Company, New Orleans, Louisiana. She is also a workforce development consultant for business, industry, and government. Contact information: 12618 Bankston Road, Amite, LA 70422. Telephone: (985) 747–0340, hwllms@bellsouth.net

M. KHATA JABOR is a member of the Faculty of Education, M. Khata Jabor, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia, khata.jabor@gmail.com