International Journal of Science and Technology
(STECH)
Bahir Dar- Ethiopia
Vol. 5 (2), S/No12, October, 2016: 24-36
ISSN: 2225-8590 (Print) ISSN 2227-5452 (Online)
DOI: http://dx.doi.org/10.4314/stech.v5i2.3

Research, Statistics and Mathematics Educators in Nigeria: Effect Size Perspective

Adeniji, K. A.
Department of Science Education
Federal University, Dutsin-Ma, Katsina State
E-mail: xyoriented@gmail.com
Tel: +2348037660561

Abstract

Over reliance on the perspective of a dichotomous reject or fail-to-reject outcome from a null hypothesis testing framework to answer research questions has become a worrisome issue to research methodologists and statistics experts. Thus, the Journals of Mathematical Association of Nigeria, Abacus (2013 & 2014) were surveyed to investigate the effect size reports practice among Mathematics Educators in Nigeria. The study showed that majority (60%) of questions of research interests of Mathematics Educators were answered by Null Hypothesis Significant Testing (NHST) and less than one –fifth (16%) of this empirical studies reported Effect sizes (standardized) for their findings. However, the survey further revealed that though Effect sizes were mostly not reported but can be conveniently estimated from the associated descriptive statistics reported by the researchers except in the analyses which seek for relationship in categorical data. Recommendations made included that Editorial Policies and guidelines of Journals in Education, especially in Mathematics Education should include Effect sizes and Confidence Intervals reports for authors.

Key Words: Research, Statistics, Effect size, Mathematics Educators, Empirical studies

Background

Starting with this fundamental question raised and answered by Kelly and preacher (2012): How do we learn from data? Simply answered it is "understanding how different statistics can be used to estimate the magnitude of a phenomenon, where the magnitude helps to address a question of interest" (p.149). Asking this thought provoking question shows how important data and statistics- a branch of mathematics are so much important in finding solution to problems facing our societies. Research as means of communicating the extent to which phenomenon occur and provide possible solutions of a felt phenomenon to policy formulators, implementers, and end users such as government and organizations, has greatly contributed to the development of this modern society. No wonder research activity is one of the integral parts of academics all over the globe.

However, over reliance on the Null Hypothesis Significance Test (NHST) by researchers to communicate their findings to their targeted audiences has been historically and continuously receiving attack from different research methodologists, Psychologists and Statistics experts because of what can be referred to as flaws or pitfalls in NHST. (Yates, 1951; Meehl, 1978; Cohen, 1990, 1994; Field, 2005; Kelly & preacher, 2012; Fritz, Morris & Richler, 2012) The following quotations buttress the point:

> The most universal reliance on merely refuting the null hypothesis is a terrible mistake, is basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology (Meehl, 1978, p.817).
>
> I have learned and taught that the primary product of a research inquiry is one or more measures of effect size, not *p* values (Cohen, 1990, p.1310)
>
> NHST: I resisted the temptation to call it Statistical Hypothesis Inference Testing (Cohen 1994, p. 997).
>
> Our view is that transitioning to a research literature focused on interval estimates of effects sizes that address the question of interest should be a top priority (Kelly & Preacher, 2012, p. 149).

Furthermore, Field (2005) raised five heart breaking reasons why reporting NHST alone to communicate the results of a finding as flaw: (1) NHST is misunderstood (2) Null hypothesis is never true (3) NHST depends on sample size (4) NHST is illogical and (5) $P< 0.05$ is completely arbitrary. Because researchers paid too much attention to the results of NHSTs and too little attention to the magnitudes of the effect size in which they are interested in, recommendations have been coming from difference organizations, editors of several journals and the American Psychological

Association (APA) to researchers to always present effect size when reporting their findings (Kelly & Preacher, 2012). For example, the Publication Manual of APA (APA, 2010) asserts that NHSTs are "but a starting point and that additional reporting element such as effects sizes, confidence interval, and extensive description are needed to convey the most complete meaning of the results" (p.33). The APA Manual further states that "complete reporting for all tested hypotheses and estimates of appropriate effect sizes and confidence intervals are the minimum expectations for all APA journals" (p.33).

Apart from the APA, many Editors of several Journals in various fields have shown their commitment into reporting effect sizes in their policies (Trusty, Thompson & Petrocelli, 2004; Thomson, 1994; Association for Psychology Science, 2011). For example, Educational and Psychological Measurement states that "authors reporting statistical significances will be required to report and interpret effects sizes" (Thomson,1994, p.845). And the International Committee of Medical Journal Editor (2007) also indicates their preference for the use of effect sizes and confidence intervals as follows:

> When possible, quantify findings and present them with appropriate indicator of measurement error or uncertainly (such as confidence intervals). Avoid relying solely on statistical hypothesis testing. Such as the use of P values, which fails to convey important information about effect size (section IV.A.6.c).

Unarguably, there is growing in attention given to the importance of reporting effects sizes and their corresponding confidence interval estimates in empirical studies. However, there is still clear missing gap between effects sizes estimates concepts and the curricular system in higher institutions. This lapse is not only peculiar to African institutions but also to some so developed nations. To buttress this, a lecturer when introducing his students (post graduates) to the concepts of effect size said: "NHST is used throughout psychology (and most other sciences) and is what you have been taught for the past 18 months (and for many Psychology undergraduates it is all they are ever taught.)" (Field, 2005, p.1). This implies that many undergraduate students (even some post graduates) in many countries in psychology and education related disciplines would or have graduated without getting familiar with the importance of effect size and confidence interval in communicating the findings of studies.

More so, the lack of encouragement from many journal editorials to report the magnitude of the effect of findings submitted for publication also impede its recognition among researchers. It has been suggested that changes in editorial policies will be required before reporting effect size will become a matter of routine (Fiddler, et al., 2005; Vacha- Haase, Nilsson, Rectz, Lance & Thompson, 2000). Now, why is

so much emphasis on effect size in place or supplement to hypothesis testing? May be because:

> Effect sizes allow researchers to move away from the simple identification of statistical significance toward a more generally interpretable, quantitative description of the size of an effect. They provide a description of the size of observed effects that is independent of the possibly misleading influences of a sample size. ...effect size can also allow the comparison of effects in a single study and across studies in either formal or informal mata-analysis (Fritz, et al., 2012, p. 2).

Many authors have given different definitions of effect size. Some authors have linking effect size with hypothesis testing, p-value, sample and some even linking it with population. But Kelly and Preacher (2012) give a well encompassing definition of effect size as a "quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest" (p.140). The two critical aspects of this definition are: (1) *Relating effect size with question of interest*- this might refer to any descriptive statistics such as central tendency, variability, association, differences, odds rate, duration, discrepancy, proportionality, superiority or degree of fit or misfit among other; (2) *Not linking the definition of effect size to a null hypothesis or NHST.* This is because each represents two fundamentally different ways of using data (Kelly & Preacher, 2012). Thus, effect size is using data to obtain a magnitude of phenomenon that addresses a question of interest while hypothesis testing or NHST is using data to obtain the probability of observing results as extreme (or more) as observed if the null hypothesis is true.

Generally, effect size has three facets in the literature. First, *Effect size measures*: These are the equations/algorithms that define or estimate particular implementation of the effect size dimension(s) of interest. These measures could be standardized descriptive statistics (e.g. Cohen's *d*; Eta squared, η^2; Correlation coefficient, *r*; Cramer's *V* etc.) or unstandardized descriptive statistics (e.g. Mean difference; Percentage, %; Bar chat; Difference in proportion etc.). Standardized measures of effect sizes are scale-free because they defined in terms of variability of the data while unstandardized measures are defined without considering the variability (Kampenes, Dyba, Hannay, & Sjoberg, 2007; Fritz et al., 2012; Kelly & Preach, 2012). In stressing the importance of the formal over the later, Fritz et al., (2012) asserted that calculating effect sizes without considering the variability in the distribution "could be seriously misleading" (p.3). Kampenes et al., (2007) further noted that because of ease in interpreting unstandardized than standardized effect size, researcher tend to always interpret the former as practically important when they are not. Though Kampenes et al., (2007) strongly recommend reports of both effect sizes to convey a meaningful research findings and for others researchers to calculate their own choice of effect sizes.

Calculations and other procedures of effects sizes are treated in many special texts. (e.g Fraenkel & Wallen 2006; Ellis, 2010; Grissom & Kim, 2011; Cumming, 2012).

Second, *Effect size values:* These are actual values calculated from certain measures. They are literally the magnitude of some phenomenon as discerned from the data, statistics or parameter. In short is a real number result obtained when effect size measure is applied to data, statistics or parameters. For instance, when examining the differences between two conditions, effect size value obtained based on "standardized differences between the means" in a study might be g= 0.8. (i.e according to Hedge's g). This value is a real number that results from applying data, statistics or parameters to an effect size measure,

$$g = \frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}}$$

where $\bar{X}_1$ and $\bar{X}_2$ donate the two means and S_{pooled} is the square root of the unbiased estimate of the within group variance. The value 0.8 implies that there is up to 71% difference between the two conditions under investigation in favour of the condition with higher mean score.

And third, *Effect size interpretation:* This simply represents the relevant interpretation of an estimated magnitude to address the information of interest. This is an abstraction related to the degree to which values differ, associated, superior over one another and so on. Thus, it is an interpretation which provides the general ideals of the way in which the question of interest will be addressed. Though some expects have made widely accepted attempts to associate regions of effects size metrics or value with descriptive adjectives terms such as "small", "medium", and "large"; "trivial", "theoretical" and "practical" and the likes to convey the meaningful effect of their findings. It has been stressed by others that the interpretation should be done in line with "area of study, research design, population of interest and research goal" (Kelly & preacher, 2012, p.146). However, some approaches have been suggested by Kampenes et al., (2007) for effect size interpretations:

(a) Standardized effect size can be interpreted in terms of the measures

(b) Standardized effect size can be compared with:

- effect sizes reported in similar experiments
- effect sizes reported in the research field in question,
- standard conventions for "small"; "medium", and "large" effect sizes developed for research in behavioural science.

Obviously, the effect size obtained from sample is in most situation not researchers' primary interest but population effect size value (Fidler & Thompson, 2001; Bird,

2002; Field, 2005; Kelly, 2008). Because we do not have access to this population effect size value, we employ the value estimated from sample to estimate the likely size of the effect in the population. A cogent reason why effect sizes and its corresponding confidence interval estimates are strongly recommended by various authors and experts (e.g Fraenkel & Wallen, 2006; Kampenes et al., 2007; Fritz et al., 2012; Kelly & Preacher, 2012) when presenting the results of studies.

In conclusion, calculating confidence interval for unstandardized effect sizes is simple and even made more easy by most statistical packages/software but calculating confidence intervals for standardized effect size is not only relatively new, complicated but also somewhat missing in almost Educational related disciplines curricular. Sequel to all these reviews and background about effect size, there is a need to investigate Nigerian mathematics educator researchers' disposition to effect size usage to communicate the meaning of their research findings.

Statement of the Problem

The issue of how data and statistics are used to solve societal problems and enhance national development through research cannot be over emphasized. But over reliance on the perspective of a dichotomous reject or fail-to-reject outcome from a null hypothesis testing framework has become a worrisome issue to research methodologists and statistics experts (Cohen, 1990, 1994; Field, 2005; Fraenkel & Wallen, 2006; Kelly & Preacher, 2012; Fritz, Morris & Richler, 2012). Despite that there is a clear push to widely report and use of effect size (ES) as a basis for communicating result and discussing the importance of those results from research studies, no study investigates the disposition of Nigerian researchers, especially in mathematics education, to its reporting and usage. Therefore, this present study looks at the extent to which mathematics educator researchers in Nigeria employ effect size reports in their studies through surveying the journal of Mathematical Association of Nigeria, *Abacus* (Mathematics Education Series) the most recognized journal of mathematics educators in Nigeria.

Objectives of the Study

The following objectives guided the study.

1. To investigate the most frequently used statistical analyses by Mathematics Educators when reporting their research findings.
2. To investigate the level of effect sizes reports among Mathematics Educators.
3. To investigate the level of report of descriptive statistics associated with Effect sizes among mathematics educators.

Research Questions

The study will answer the following questions:

1. What are the most frequently used statistical analyses by the mathematics educators in Nigeria to report their research findings?
2. To what extent do mathematics educators in Nigeria report effect size estimate for their studies?
3. To what extent do mathematics educators in Nigeria report descriptive statistics associated with effect sizes for their studies?

Methodology

The study employed descriptive research design approach. The population for the study comprised all the mathematics educator researchers in Nigeria and a purposive sampling techniques was employed to select their journal Articles from the *Abacus*- Journal of Mathematical Association of Nigeria (Mathematics Education Series). Forty-eight (48) Articles (excluding 19 theoretical and 3 empirical) were surveyed from the two most recent editions of the *Abacus* (2013 & 2014). Each finding was categorized under statistical tools used to address them as question of interest; Inferential, Descriptive and Both inferential and descriptive. Based on the purpose of the study, those questions addressed with magnitude of difference/relationship (Effect sizes) were further categorized into Standardized, Unstandardized and Both. Finally, all the empirical studies were surveyed to verify the level of associated descriptive statistics reports for easy effect sizes (standardized) calculations.

Results

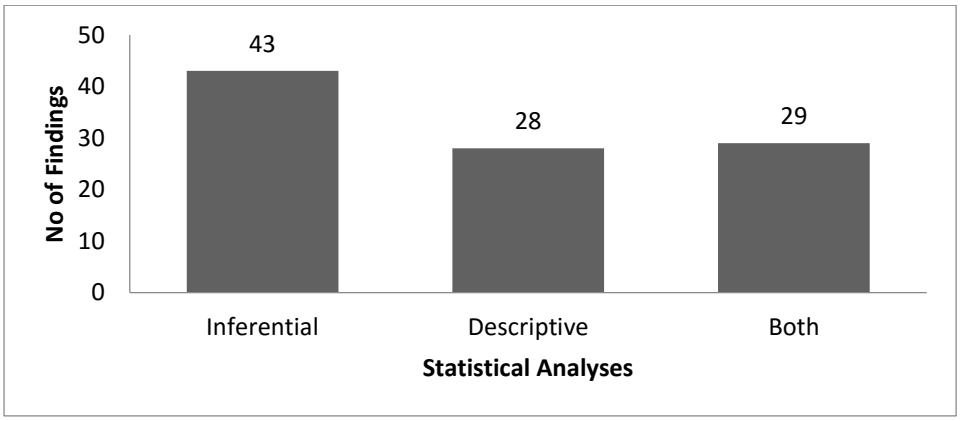

Figure1: Level and type of statistical analyses employed by mathematics Educators

Figure 1 indicates that 43 of all research findings reported by mathematics education in Nigerian were done using inferential analyses (NHST). Analyses in this group included analyses of (co)variance, *t*-test, correlation and chi-square with AN(C)OVA and *t*-test reported more frequently. Also, 29 of their findings were reported using both inferential and descriptive analyses. And all those findings (28)

reported only with descriptive analysis were not carried out to compare/relate two or more variables. They were presented only to describe the condition(s) being studies. These included means, percentages, % and Charts. Therefore, majority (60%) of empirical findings among mathematics educator researchers in Nigeria are reported mostly by the NHST.

Figure 2: Level and type of effect sizes reporting among Mathematics Educators

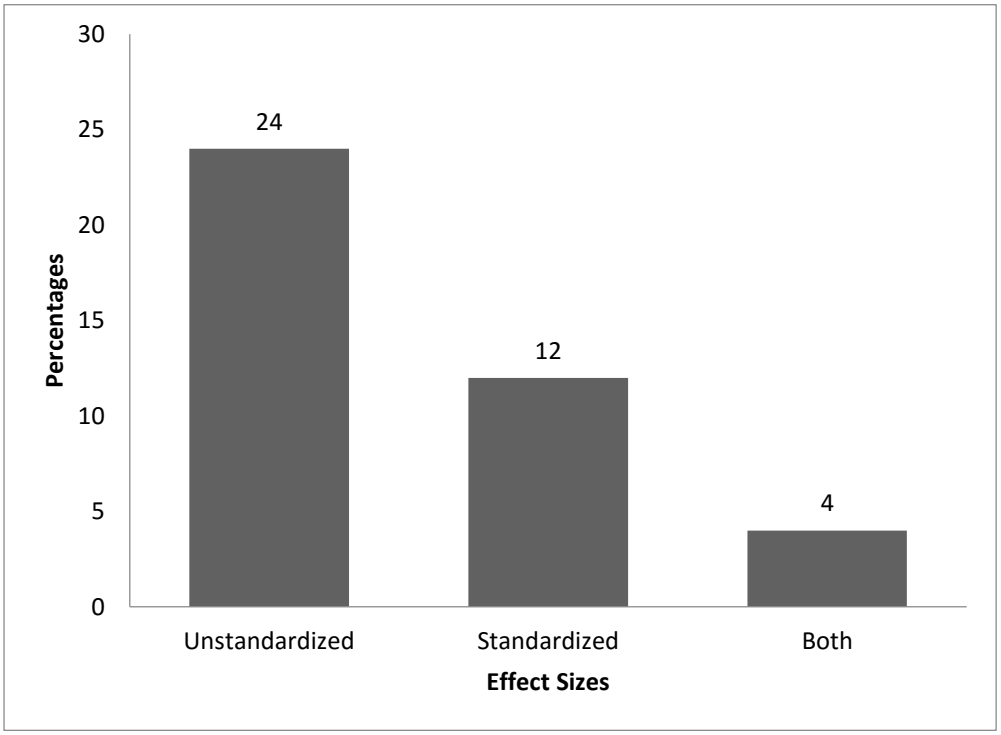

Figure 2 shows that 17(24%) out of 29(40%) research findings that reported magnitude of their findings by mathematics education in the country employed unstandardized effect size and 9(12%) employed standardized effect size while 3(4%) used both. It was observed in the analysis that all the findings 12(9+3; 16%) that reported standardized effect size came from ANOVA/ANCOVA tests. The reason for this is very close to the fact that SPSS software produces these estimates not because researchers really wanted it. As they were not used to present the magnitudes of their findings when interpreting the results.

Figure 3: Level of reporting descriptive statistics associated with effect sizes estimations

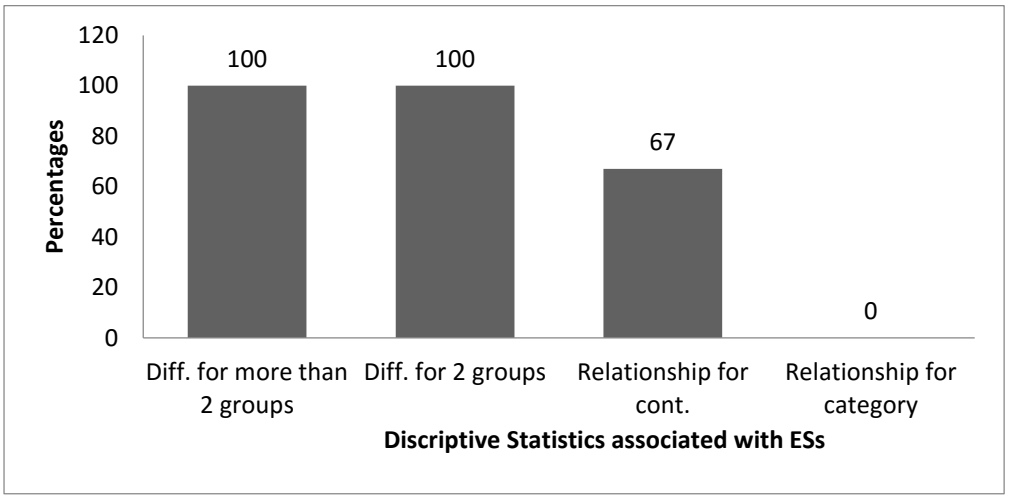

Figure 3 depicts that all associated descriptive statistics for easy calculation of effects size (standardized) were fully reported for analysis which investigate the differences between two or more variables. These included *Sum of Square(factors*) and *Sum of Square(total)* or *N*, *means* and *Standard Deviations* for each group. A reasonable amount (67%) empirical studies which investigated relationship between variables for continuous data reported full associated descriptive for calculating standardized effect size while no single analysis reported required Descriptive Statistics to estimate effect size for question which investigated relationship between variables for categorical data.

Discussion

The quest to find out the practice of effect size reports among Mathematics Educators in Nigeria in this present study revealed that majority of these researchers still heavily rely on NHST to proffer solutions to the questions of interest in their studies. This revelation supports the comments of Meehl (1978) and more recently by Fritz et al., (2012). Also, it was observed that most of Nigerian researchers have not recognized effect sizes estimates as tool to present the magnitude of their findings. This finding is in perfect agreement with Kampenes et al., (2007) position which noted that only 29% experiments reported effect sizes (un-and standardized) in software engineering studies and that of Fritz et al., (2012) which submitted that far less than half of analyses reported effect sizes (standardized) in Psychology journal surveyed. This assertion was

made from the study because those analyses (16%) that reported effect sizes did not use them in interpreting their findings.

However, information provided by the researchers in mathematics education was sufficiently enough to calculate effect size for any test comparing two or more conditions. This observation contradicts that of Kampenes et al., (2007) and Fritz et al., (2012) which observed that descriptive statistics associated with the test comparing more than two conditions were sparsely reported but supports latter for the test comparing two conditions. The study further observed, in contrary to Fritz et al., (2012) that a tangible number of results reporting relationship between variables for continuous data in the *Abacus* reported descriptive statistics which allow easy calculation of effect sizes for the data.

Lastly, the reverse situation was observed in the non-parametric test which seeks to find out the relationship between two variables in categorical data. This is because no single analysis reported complete components of the analyses for smooth estimation of effect size from the data. The revelation confirmed Fritz et al., (2012) but once more contradicts that of Kampenes et al., (2007) that found about 33% reports for this non-parametric test in their study.

Conclusion

It was observed that majority of research findings among mathematics educator in Nigeria rely solely on Null Hypotheses Significance Testing. Descriptive statistics such as standardized effect sizes have not been seen as a statistical tool to answer their questions of interest as those that did (though unstandardized) to present magnitude of difference/relationship still went ahead to use NHST. It can also be concluded that less than one- fifth of the analyses conducted by these group of researchers reported effect size (standardized). However, other researchers can easily calculate effect sizes for question of interest that seek for difference in two or more variables from the associated descriptive statistics provided. Finally, tangible number of their empirical studies which seek to find the relationship between two variables presented required descriptive statistics to calculate their effect sizes while for those that seek relationship for categorical data will be difficult to calculate because of insufficient reports of these associated descriptive statistics.

Recommendations/Implications

It very important to make the following recommendations/implications based on the major findings from the study

- *Editorial Policies of Education Journals*: The editorial policies and guidelines for authors in education journals especially in mathematics should include effect size and confidence interval reports.

- *Curriculum in Education related Disciplines*: The study revealed that if mathematics educators are still not getting it fully wright about applications of statistics in research, then the concepts of effect size and confidence interval are somewhat missing in the Educational Statistics/Research Methods courses in our education related disciplines at undergraduate and possibly postgraduate levels. This situation needs quick revisit by curriculum planners.
- *Research findings interpretations*: Though not presented as part of the survey, there were clear evidences in the study that indicated overvaluing/underestimating findings of some studies as result of over relying on NHST to communicate the research results. Researchers in the field can investigate this in the future research.

Limitations

The main limitations to this survey are the overgeneralization of the findings based on the assumptions made about the *Abacus* as the most used journal of Mathematics Educators in Nigeria. The bias regarding articles and analyses selection and possible inaccuracy in data extraction due to complexity of data usage of some researchers may also be considered as another limitation.

References

American Psychological Association. (2010). *Publication manual of the American psychological association* (6th edn.). Washington, DC: Author.

Association of Psychological Science. (2011). *Submission guidelines for psychological science*. Retrieved from http://www.psychologicalscience.org/index.php/publications/journals/psychological_science/ps-submissions

Bird, K. D. (2002). Confidence intervals for effect sizes in analysis of variance. *Educational and Psychological Measurement*, 62, 197-226. doi: 10.1177/0013164402062002001

Cohen, J. (1990). Thing I have learned (so far). *American Psychologist*, 45, 1304-1312. doi: 10.1037/003-066X.45.12.1304

Cohen, J. (1994). The earth is round ($p < 0.05$). *American Psychologist*, 49, 997-1003.

Cunning, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals and meta-analysis.* New York, NY: Routledge.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis and the interpretations of research results*. Cambridge, United Kingdom: Cambridge University Press.

Fidler, F. & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed- and random- effect sizes. *Educational and Psychological Measurement*, 61, 675-604. doi:10.1177/0013164401614003

Fidler, F., Cumming, G., Thomason, N., Pannuzzo, D., Smith, J., Fyffe, P., Edmonds, H., Harrington, C. & Schmitt, R. (2005). Toward improved statistical reporting in the journal of consulting and psychology. *American Psychological Association*, 73(1), 136-143.

Field, A.P. (2005). *C8075 (Research Methods 2): Effect sizes*. Retrieved from http://www.effectsizes.pdf

Fraenkel, J.R., & Wallen, N.E. (2006). *How to design and evaluate research in education.* New York: McGraw-Hill Companies, Inc.

Fritz, C. O., Morris, P. E. & Richler, J. J. (2012). Current use, calculations and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18.

Grissom, R. J. & Kim, J. J. (2011). *Effect sizes for research: A broad practical approach* (2nd edn.). New York, NY: Psychology Press.

International Committee of Medical Journal Editors. (2007). *Uniform requirements for manuscripts submitted to biomedical Journals: Writing and editing for biomedical publications.* Retrieved from http://www.icmje.org/2007_urm.pdf

Kampenes, V. B., Dyba, T., Hannay, J. E. & Sjoberg, D. I. K. (2007). *Systematic review of effect size in software engineering experiments.* Retrieved from http://www.R04-effectsize-kampenes-ist-nov07.pdf

Kelley, K. (2008). Sample size planning for the squared multiple correlation coefficients: Accuracy in parameter estimation via narrow confidence intervals. *Multivariate Behavioral Research, 43*, 524–555. doi: 10.1080/00273170802490632

Kelly, K. & Preacher, K. J. (2012). On effect size. *Psychological Methods*, 7(2), 137-152.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Roland, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, 54, 837-847.

Trusty, J., Thompson, B. & Petrocelli, J.V. (2004). Practical guide for reporting effect size in quantitative research in the Journal of Counseling and Development. *Journal of Counseling and Development*, 82, 107-110.

Vacha-Haase, T., Nelson, J. E., Reetz, D. R., Lance, T. S. & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory and Psychology*, 10(3), 413-425.

Yates, F. (1951). The influence of statistical methods for research workers on the development of the science of statistics. *Journal of the American Statistical Association*, 46, 19-34. doi: 10.2307/2280090.