# HIGHLY SIGNIFICANT FINDINGS IN PSYCHOLOGY: A POWER AND EFFECT SIZE SURVEY ***

BY

ROSE MARIE WARD

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2002

UMI Number: 3053127



UMI®



ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

DOCTOR OF PHILOSOPHY DISSERTATION

OF

ROSE MARIE WARD

APPROVED:

Dissertation Committee:

Major Professor

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2002

Abstract

The quality of current psychological research has been questioned because of perceived flaws in the primary methods of inquiry. As early as the late 1930s researchers began criticizing the methods by which psychological research examines hypotheses. The current method of hypothesis testing represents a hybrid of two models of significance testing. The originators of the models never intended for the models to be combined and never suggested the present model. Early psychology textbooks were the first to present the model of hypotheses testing, as we know it today.

The cut and paste manner of the current model of significance testing currently exhibits fatal flaws. Utilizing the Null Hypothesis Significance Testing (NHST) method can result in two errors or a correct decision. The researcher can make the correct decision by either rejecting a false null or failing to reject a true null. The errors occur when a researcher fails to reject a false null or rejects a true null.

Over the last 50 years, researchers have suggested that the American Psychological Association require additional indices to augment findings of statistical significance. Among the indices suggested are statistical power and measures of magnitude of effect size. Statistical power is a consideration that seems to complement NHST perfectly. Statistical power is defined as the probability of discovering a statistically significant result, which should be an automatic concern for researchers utilizing NHST. Measures of effect

size estimate the magnitude of the treatment effect or the differences between groups. They answer the question, "How much of a difference is found?" Similarly, this measure seems to harmonize with the NHST procedure.

While researchers and the APA have suggested utilizing these and other methods to supplement NHST, it seems that researchers are not currently using statistical power and effect size measures to their fullest extent. The current study examined the articles in three psychology journals to assess the current status of statistical power and effect size measures. The results of the current study suggest that about 7% of studies estimate or discuss statistical power, and about 30% calculate effect size measures. These numbers are far below the desired level of mandatory reporting of these measures. Also, when statistical power was calculated for 157 articles (45 in the *Journal of Personality and Social Psychology*, 57 in the *Journal of Consulting and Clinical Psychology*, and 55 in the *Journal of Abnormal Psychology*) for 2,747 statistical tests for a total of 27,705 power calculations (power was calculated for effects beyond the normal small, medium, and large), a slight increase (above the original 1962 study and the replication in 1990) in statistical power was noted. In terms of effect size measures, a medium effect size was discovered as the average effect size across studies, which confirms previous researchers speculations about the average effect size in psychological research.

It would seem that though the average effect size in the current research is of medium size, current research designs do not have sufficient statistical power to detect such an effect size. The current research should also strive to improve current statistical power survey methods to incorporate more advanced statistical methods to gain a more representative evaluation of the average effect size in psychological research.

## Acknowledgements

If I did not believe in God, if I did not pray every day, if I did not place this dissertation in God's hands, it never would have been completed. I found the strength I needed to complete this task through God.

*Philippians 4:13 - I can do all things through Christ which strengtheneth me.*

Without my husband, John, I probably would have given up on finishing this project long before I started. He reminded me daily that I could do it, he made my lunch, he cleaned the house when I was too tired to think, he made sure I took breaks and ate, he held me when I just wanted to quit, he was my cheerleader, and he was my prayer partner. Thanks for believing in me.

I would like to thank my major professor, Dr. Joseph S. Rossi. With his guidance and support, I completed this dissertation project. Next, I would like to thank my committee members, Dr. Joseph Fava, Dr. Colleen Redding, Dr. Bryan Blissmer, Dr. Geoffrey Greene, and Dr. Sara Johnson, for their guidance and suggestions on earlier drafts. I would also like to thank Dr. Jerry Cohen for loaning me his prized Journals of Personality and Social Psychology, and Dr. Mark Wood for loaning me his Journals of Abnormal Psychology.

Finally, I would like to thank my family. They reminded me that learning and joking are a way of life. Thanks for teaching me the importance of an education and making me smile when I was too serious.

## Preface

Every student eventually becomes a teacher. Whether they are a formally employed as a teacher or a professor, or they just become a parent, all students become teachers. This dissertation taught me a lot about many different aspects of psychology, not only the aspects I was looking for, but also many underlying aspects. This dissertation is an effort to replicate earlier power surveys and extend the research on effect size measures, but in the process of being a student of the information, I found myself teaching the material to my students in my introduction to statistics class. This poem and dissertation reflects many of things that I learned from them and they learned from me:

Note to a current student from your professor:

Why do you look at me that way? I stand before you explaining themes and concepts, facts and figures. Carefully repeating items because they'll be on your exam. Trust me, I wrote it last night.

Why do you sit there as I ask a question talking to your friends, commenting about my clothes, hairstyle, mannerisms, or voice? Why can't you just listen to my message and jot down some ideas. Maybe a head nod every so often to let me know that you understand.

Do you understand? I see the doodles on your papers. I see the homework you are working on--that you think that you have hidden beneath your book.

Your book-do you even know what chapter we are on? Have you even bought the book yet? I spend hours creating a lecture to compliment the text. But in order for you to get the gist, you have to read the book.

The tests- I try to be fair. Not too many multiple choice, a couple essays. Trying to target what you need to know from this class. Don't you study or do you think that your past history as a student is going to enhance your grade?

Sometimes I don't understand you. You answer my question with little thought or no desire. Like you had been thinking of dinosaurs before I asked you and were angered by my intrusion.

I came here to teach. It is something I've always wanted to do. I can teach all day, but I need your help. I need someone on the receiving end. I need you to learn.

Note to a professor from a current student:

I know I am here to learn. I know I have responsibilities. Trust me I go home to read more often than you think. I am paying attention in class even when my head is down. Remember that as you have bad days, so do I. There will be days that I am excited to learn, but realize that you can help me with that.

So remember:

Don't read to me. I don't come to class like a kindergartner. I like to read on my own time. TALK to me. I am a person. Lectures are great, but discussions are better. It's ok to tell be the facts, but help me develop my own conclusions. Communication is the key. Use handouts, visuals, e-mail, visuals... Let me say visuals again. Nothing compliments a verbal message better than a visual.

Now the verbal message. Work hard on your verbal message. WE can tell when you decided to use the same lecture for the past 20 years. Voice inflection, hand gestures, and eye contact, basically think of every class as an interview--an interview with your students. It is an interview with us. Maybe it would help you to think of it as a test. We are testing you every class. We want to see how good you are.

Tests- test on what you taught us. Don't test us on what you think we should know. Professor, guider, teacher show me what you want me to learn. Learn, that's why I'm here. That's why you are in front of the class. When you boil down our classes can you answer -what did <u>we</u> learn today?

Yes, I said we. We are learning together.

And that is what <u>we learned</u> today.

Note to a student from your professor:

You want to learn. You want us to learn together. I am open to suggestions. But you never come to see me during my office hours. When you are struggling or uncomfortable with the material, I never know unless you tell me. I have watched you over the semester go through your moods. I have seen you beaming and I have seen the anger weld up beneath your lips.

I am not too old to remember what it was like to be a student. I remember all of the pressure and frustration you feel. But for me to help you better we need to communicate. I need you to be involved. When I try new things, new lectures, new exercises, I need to hear what you think so that I can improve too.

For us to learn together, we both need to have active roles. I need to see your eyes when we are talking.

Note to a professor from your student:

I don't pay attention all of the time. I am sorry. Maybe I should volunteer an answer every so often. Maybe I should show you that I have read the readings. But I need you to be talking to me and not just talking to hear yourself speak.

If I don't' feel your warmth, I am not going to come to your office hours. If I do come, I feel like I am admitting that I am not capable. If I need help, then something must be wrong.

Do you see me trying to improve my grade? Every so often, I take a couple more notes then I used too. If the class is struggling as a whole with the material, do you blame us for not studying or acknowledge that the test may have been unfair?

I read your comments and know that they are valid. I really think that we need to work together to improve them.

A note from a professor to a current student:

When I look out in my classes sometimes, I see children who have been spoon fed their whole lives. Teachers gave you the answers so that you could pass the test. All your life you have thought that you were entitled to the world. Never did you once consider working for it.

You think that my tests are too hard? Or maybe unfair? Or that my lectures are boring? Do you realize that I am trying to challenge you to think?

So I give a lot of readings. So my tests are hard. I am not here to get high student ratings at the end of the semester. I am here to give you knowledge. I am here to help you into your future.

A note from a current student to a professor:

My future does not seem real yet. It is hard for me to think about a future when I don't even know what I am doing tonight.

Don't you think that I can give you knowledge too? Don't you think that you could learn from me too?

A note from a professor to a current student:

Yes, I can learn from you, but you don't seem interested.

A note from a current student to a professor:

I can learn from you too, but you don't seem interested.

This dissertation is an effort to walk in the steps of my predecessors. Recently, my major professor and I tried to trace my psychological lineage. As far as we know it is as follows (from recent to past):

Joseph S. Rossi (University of Rhode Island, Ph.D., 1984)

Charles E. Collyer (Princeton University, Ph.D., 1976)

Ronald A. Kinchla (University of California, Ph.D., 1962)

Richard C. Atkinson (University of California, Ph.D., 1950?)

William K. Estes (University of Minnesota, Ph.D., 1943)

Burrhus Frederic Skinner (Harvard, Ph.D., 1931)

Edwin Garrigues Boring (Harvard, Ph.D, 1908?)

Edward Bradford Titchener (University of Leipzig, Ph.D., 1892)

Wilhelm Max Wundt (University of Heidelberg, M.D., 1856)

It is my hope that one day, I will become a teacher like those who came before me.

April, 2002 *Rose Marie Ward*

## Dedication

I dedicated this to my family.

*My husband, John*

*Raymond and Mary Regan*
*Ray and Sara Regan*
*Dan and Lisa Regan*
*Michael Regan*
*Bernadette Regan*
*John Paul Regan*
*Thomas Regan*
*Robert Regan*
*Elizabeth Regan*

*Willie and Billie Ward*
*Willie and Satomi Ward*
*Caddell and April Bachelor*
*Eve Bachelor*
*Gigi Ward*
*Laura Ward*

Without them, I would have never developed this wacky sense of humor.

Table of Contents

List of Tables

List of Figures

*Delight yourself also in the Lord,*
*And He will give you the desires of your heart.*
*Commit your way to the Lord,*
*Trust also in Him*
*And He shall bring it to pass.*
*Psalms 37:4-5*

## Chapter 1:

## Statistical Power and Effect Size: Why Bother?

For years, the American Psychological Association (APA) has been "encouraging" psychologists to calculate power prior to running their experiments and to calculate effect sizes for their statistics when the studies are complete (most recently APA, 2001). The APA's push for power calculations and effect size indices resulted from numerous researchers publishing the flaws and problems with current statistical methods (e.g. Hunter, 1997; Kirk, 1996; Harlow, Mulaik, & Steiger, 1997; Morrison & Hinkle, 1970; Schmidt, 1996; Thompson, 1996, 1997, 1999). Statistical power estimates and effect size indices have been presented as solutions to the current problems (power: Cohen, 1962; effect size: Thompson, 1999).

### Definition of Statistical Power

In research, there are four possible outcomes to a study. One possible outcome is that in the population the treatment effect does not exist and the researcher rejects the null hypothesis. In this case, the researcher has made a Type I error. Another possible outcome is that the effect is not present and the researcher fails to reject the null hypothesis. This decision is the correct statistical decision. Another possible outcome is that the effect is present and the researcher fails to reject the null hypothesis. This decision is known as a Type II Error. The final possible outcome is when the effect is present and the researcher correctly rejects the null hypothesis. This is known as the power of the statistical test. (See table 1.1 for the table of statistical possibilities.)

---

Insert Table 1.1 here

---

Cohen (1988) defined power of a statistical test as "the probability that it will yield significant results." It is important to note the power discussed in this study refers to power calculated a priori. Statistical power is reliant on three factors: significance level, sample size, and effect size.

Researchers have often emphasized how statistical power and null hypothesis significance testing should be coupled together (Cohen, 1988; Rossi, 1990). Given that statistical power is the likelihood of obtaining a significant result, one would assume that researchers would seek to examine the statistical power of their study a priori.

Determining Statistical Power

In order to operationalize statistical power, researchers have to make assumptions about the data. Many researchers have utilized the standard definitions set forth by Cohen (1977) and developed statistical programs for calculating power (i.e., PASS: Hintze, 2001). These programs use Cohen's assumptions about effect size to ease the calculation of statistical power. The current study employed Rossi's (1984) BASIC programs and PASS (Hintze, 2001) designed for the same purpose.

Effect Size Measures Used to Calculate Power

Power determinations were made using Cohen's (1962, 1977) definitions of small, medium, and large effect size. Below are the primary effect sizes for which power was calculated which are adapted from Cohen's (1988) power textbook. The definitions for small, medium, and large effect sizes for each statistic are listed in table 1.2 and are based on Cohen's 1962 and 1977 definitions.

---

Insert Table 1.2 here

---

1. **t-test**: Cohen suggested using a standard metric for the student's t-test. He developed a metric dubbed Cohen's d, which represents the standardized difference between group means (Cohen, 1988):

(1.1) $$d = \frac{(M_1 - M_2)}{s}$$

where $M_1$ is the mean of the first group, $M_2$ is the mean of the second group, and s is pooled standard deviation. Cohen (1988) described the relationship of $d$ to delta, the noncentrality parameter (NCP) for the noncentral t distribution, as follows:

(1.2) $$\delta = d\sqrt{(n/2)}$$

where $n$ is the sample size for each group.

2. **Pearson r:** The effect size measure for Pearson r is the correlation coefficient.

3. **Differences between correlation coefficients:** Cohen (1969) developed an effect size index for the difference between correlations, $q$. This index utilizes a transformation of Fisher's $r$ to $z$ scores:

(1.3) $$q = |z_1 - z_2|$$

To transform $r$ to a $z$ score:

(1.4) $$z = \frac{\ln((1+r)/(1-r))}{2}$$

An equivalent form of the above formula is presented below:

(1.5) $$z = \text{arctanh}\,(r)$$

Rossi (1985) published tables for computing $q$.

5. **Differences between proportions:** Cohen (1988) developed a method for calculating the difference between proportions using $\phi_1$ and $\phi_2$, which are the arcsine transformation for the two proportions.

(1.6) $$h = |\ \phi_1 - \phi_2\ |$$

Eisenhart (1947) suggested using the arcsine transformation to stabilize the variance and normalize the distribution of proportions:

(1.7) $$\phi = 2 \text{ arcsine } \sqrt{p}$$

6. **Chi-square tests:** Cohen also developed a standardized effect size measure for the chi-square test. He called it $w$:

(1.8) $$w = \sqrt{(\Sigma((P_{1i} - P_{0i})^2/P_{0i})}$$

where $P_{0i}$ is the proportion in cell $i$ specified by the null hypothesis, $P_{1i}$ is the proportion specified by the alternative hypothesis, and they are summed from 1 to m (m= the number of cells). He described the relationship between $w$ and $\lambda$ (the NCP of the noncentral chi-square distribution) as:

(1.9) $$\lambda = w^2 * N,$$

where $N$ is the total sample size.

7. **F tests in the analysis of variance**: Cohen defined the effect size measure for the ANOVA as $f$:

(1.10) $$f = \frac{s(m)}{s}$$

where $s$ is the pooled standard deviation of the $k$ groups, and $s(m)$ is the standard deviation of the $k$ groups. For the two-group case, $f$ is related to $d$, the effect size index for the $t$ test, by

(1.11) $$f = \frac{d}{2}$$

The index $f$ is also closely related to $\varphi$, the NCP of the noncentral f distribution introduced by Tang (1938):

(1.12) $$\varphi = f * \sqrt{(n)},$$

$f$ is also related to $\lambda$, the NCP used by Patnaik (1949):

(1.13) $$\lambda = f^2 * n * k.$$

8. **F tests in multiple regression/correlation analysis**: Cohen (1977) suggested $f^2$ as a standardized measure of effect size for multiple regression and correlation analysis:

(1.14) $$f^2 = \frac{R^2}{(1 - R^2)}$$

where $R^2$ represents the squared multiple correlation coefficient. It represents the proportion of variance in the dependent variable accounted for by the set of independent predictor variables. It is related to λ, the NCP for the noncentral *F* distribution, as follows:

(1.15) $$\lambda = f^2 * v,$$

where

(1.16) $$v = N - k - 1.$$

In the above formula, $v$ represents the degrees of freedom for the error term, $N$ is the total sample size, and $k$ is the number of groups.

9. **F test for the one-way multivariate analysis of variance:** Cohen (1988) used $f^2$ as the measure of effect size for MANOVA. F has a slightly different definition when being applied to MANOVA:

(1.17) $$f^2 = L^{-1/S} - 1,$$

where

(1.18) $$L = |E| / |E + H|,$$

Here, $L$ = Wilks' λ, E is an error matrix, and H is an hypothesis matrix and

(1.19) $$S = \frac{\sqrt{(k^2_Y k^2_X - 4)}}{(k^2_Y + k^2_X - 5)}$$

where $k_Y$ and $k_X$ are the numerator and denominator degrees of freedom.

Rossi (1998) developed BASIC software to accomplish this type of analysis. This program was used for computing the power of one-way MANOVA in this study.

10. **Covariance Structure Modeling:** MacCallum, Browne, and Sugawara (1996) developed a framework to assess the power of covariance structure modeling or structure equation modeling (SEM) utilizing the Root Mean Square Error of Approximation (RMSEA). Given the values for alpha, degrees of freedom, sample size, the null hypothesis value of RMSEA, and the alternative hypothesis value of RMSEA, the power for a SEM can be calculated. For a complete discussion of the formulas involved, refer to MacCallum et al. (1996).

11. **Meta-Analysis:** Hedges and Pigott (2001) devised a method for assessing the power of a meta-analytic survey on the basis of whether the survey was a fixed- or random-effects test. They explored the power of meta-analytic surveys assuming that the effect sizes presented were for *k* independent studies and that the conditional variance is not known.

Fixed Effect Surveys: When examining a survey that address mean differences, the following equations are appropriate:

(1.20) $$d_i = (Y_{Ai} - Y_{Bi})/s_i$$

The above formula defines the standardized mean differences (Hedges & Olkin, 1985) where $Y_{Ai}$ and $Y_{Bi}$ are the sample means of the two samples of interest for study $i$ and $s_i$ is the pooled sample standard deviation for each study. To estimate the variance of $d_i$, the following formula is used:

(1.21) $$v_i = \frac{n_{Ai} + n_{Bi}}{n_{Ai}\, n_{Bi}} + \frac{d^2_i}{2(n_{Ai} + n_{Bi})}$$

where $n_{Ai}$ and $n_{Bi}$ are the sample sizes of the two samples of interest in the $i$th study. Then the value of the weighted mean effect size is:

(1.22) $$v. = v/k$$

where $k$ is the number of studies. The next step is to calculate lambda using formula 1.23.

(1.23) $$\lambda = (\theta - \theta_o)/\sqrt{(v)}$$

$\theta$ represents the population effect size which is estimated and lambda is the mean of a normal distribution with a variance of 1. The power of a 1-tailed test is:

(1.24) $$p = 1 - \Phi\,(c_\alpha - \lambda)$$

where $\Phi$ (x) is the standard normal cumulative distribution function. A two-tailed test has the following formula for power:

(1.25) $$p = 1 - \Phi\,(c_{\alpha/2} - \lambda) + \Phi\,(-c_{\alpha/2} - \lambda)$$

For surveys that examine correlation coefficients and tests of heterogeneity of effect size parameters, similar formulas are used (refer to Hedges and Pigott, 2001, for a complete discussion).

Random Effect Surveys: When examining surveys that address mean differences, the following power equations are appropriate. For a one-tailed test:

(1.26) $$p = 1 - \Phi\,(c_{\alpha/2} - \lambda^*)$$

and the following for a two tailed test:

(1.27) $$p = 1 - \Phi\,(c_{\alpha/2} - \lambda^*) + \Phi\,(-c_{\alpha/2} - \lambda^*)$$

where lambda is defined as:

(1.28) $$\lambda^* = (\mu - \mu_0)/\sqrt{(v^*)}$$

Hedges and Pigott (2001) discuss the random effects procedures in complete detail.

Definition of Effect Size

The term "effect size" has come to represent a family of indices that measure the magnitude of an experimental effect or how effective the treatment was. Unlike significance tests and power, the effect size index is not influenced by sample size. Researchers control the size of the sample and the alpha level at which they will test, but they do not control the effect size.

Other researchers have commented on the aspects of research which influence effect size. "The effect size obtained in a research study depends on a variety of factors, including (a) the potency of the treatment, (b) the reliability and validity of the outcome measures in relation to the treatment, and (c) the amount of uncontrolled variation in the research design"

(Kosciulek & Szymanski, 1993, p. 213). While a researcher cannot directly influence effect size, they can make efforts to ensure that their research gives the best estimate of the effect size possible by reducing error and using reliable, valid measures.

<u>What is an Effect Size?</u>

Simply, an effect size is the magnitude of the treatment effect:

(1.29) $$Pr(X_1 > X_2)$$

It is the probability that a random sample for population 1 will be greater than a random sample from population 2. It is measured in two ways. First, it can be estimated as the standardized difference between two means. Secondly, it can be considered the correlation between the independent variable classification and the individual scores on the dependent variable (Rosnow & Rosenthal, 1996).

<u>Two Independent Groups</u>

When examining two independent groups, one can estimate the magnitude of the difference between groups using a variety of effect size measures. The primary measure discussed in this case is Cohen's d. It is a descriptive measure that can be calculated from the following formula:

(1.30) $$d = \frac{M_1 - M_2}{\sigma}$$

Where σ represents pooled standard deviation and can be calculated from the following:

(1.31) $$\sigma = \sqrt{\frac{\Sigma(X - M)^2}{N}}$$

(Note: Formula 1.30 and formula 1.1 are the same.) X, M, and N represent the raw score, mean, and number of cases respectively. Cohen's d can also be computed from the value of student's t:

(1.32) $$d = \frac{2t}{\sqrt{(df)}}$$

Formula 1.32 assumes equal sample sizes. If the groups are unequal, the following formula is appropriate:

(1.33) $$d = \frac{t\,(n_1 + n_2)}{\sqrt{(df)}\sqrt{(n_1 n_2)}}$$

Cohen's d can also be calculated from Pearson's correlation, *r*:

(1.34) $$d = \frac{2r}{\sqrt{(1-r^2)}}$$

If given Hedge's *g*, using the following formula, one can transform Hedge's *g* to Cohen's *d*:

(1.35) $$d = g\sqrt{(N/df)}$$

As mentioned above, Cohen (1988) operationalized the definition of his standardized measure of effect size into small, medium, and large effects. This was an attempt, based on his research experience, to define effect sizes and categorize them into metrics. Table 1.3 presents Cohen's definitions (small, medium, and large), effect sizes, and the percent of nonoverlap between the treatment group and the control group. Cohen defined small as an effect size of 0.2 which means the 14.7% of the two distributions of interest do not overlap whereas a large effect size, 0.8, has 47.7% of the distributions not overlapping.

---

Insert Table 1.3 here

---

These effect sizes can be thought of as an average percentile standing. When the data are examined, the researcher wants to know how different is the average percentile standing of the average treatment group participant in comparison to the average untreated participant. An effect size of 0.8 (Cohen defines as large), means that the treatment group is at the $79^{th}$ percentile in comparison to the control group. If there is an effect size of zero, the mean of the treatment group is the $50^{th}$ percentile of the control group.

Another effect size measure for two independent groups is Hedge's g. It was named for the pioneer of meta-analysis, Gene V Glass. It is defined as:

(1.36) $$g = \frac{M_1 - M_2}{\sigma}$$

Where σ represents pooled standard deviation and can be calculated from the following:

(1.37) $$\sigma = \sqrt{\frac{\Sigma(X - M)^2}{N - 1}}$$

X, M, and N represent the raw score, mean, and number of cases respectively. Pooled standard deviation can also be calculated from the following formula:

(1.38) $$\sigma = \sqrt{(MS_{within})}$$

Hedge's g can also be computed from the value of student's t:

(1.39) $$g = \frac{t\sqrt{(n_1 + n_2)}}{\sqrt{(n_1 + n_2)}}$$

Formula 1.31 assumes equal sample sizes. If the groups are equal, the following formula is appropriate:

(1.40) $$g = \frac{2t}{\sqrt{(N)}}$$

Similar to Cohen's *d*, Hedge's *g* can also be calculated from Pearson's correlation, *r*:

(1.41) $$g = \frac{\frac{r}{\sqrt{(1 - r^2)}}}{\sqrt{(df(n_1 + n_2))/(n_1 n_2))}}$$

If given Cohen's *d*, using the following formula, one can transform Cohen's *d* to Hedge's *g*.

(1.42) $$g = \frac{d}{\sqrt{(N/df)}}$$

Correlational Measures of Effect Size

The effect size correlation can be defined as the relationship between a dichotomous independent variable and a continuous dependent variable.

(1.43) $$r = r_{dv,iv}$$

It can be computed from the students' *t* value:

(1.44) $$r = \sqrt{(t^2/(t^2 + df))}$$

It can also be computed from a chi-square test result if the chi-square has one degree of freedom.

(1.45) $$r = \Phi = \sqrt{(\chi^2/N)}$$

This value is also known as phi, Φ. If the researcher performs an *F* test with two groups (a single degree of freedom *F* test), the correlation effect size can be computed using the following formula:

(1.46) $$r = \sqrt{(F/(F + df_{error}))}$$

The effect size correlation also has a fairly simple relationship with Cohen's d:

(1.47) $$r = d/\sqrt{(d^2+4)}$$

To compute it from Hedge's *g* is a little more complicated:

(1.48) $$r = \sqrt{\{(g^2 n_1 n_2)/[\,(g^2 n_1 n_2)+(\,g^2 n_1+n_2)df]\}}$$

Cohen also operationalized definitions for small, medium, and large effect sizes for the correlation. He chose 0.1, 0.3, and 0.5 respectively. Table 1.4 compares Cohen's definitions for small, medium, and large for Cohen's *d*, *r*, and percent of variance accounted for ($r^2$). Using the formulas provided above (adapted from Cohen (1988) and Rosnow and Rosenthal (1996)), the values in table 1.4 were calculated. A large effect in correlational research accounts for 13.8% of the variance in the dependent variable from the independent variable. Small and medium effects account for 1.0 and 5.9 percent respectively.

---

Insert Table 1.4 here

---

Effect Size Measures for the ANOVA

Effect size measures for the analysis of variance (ANOVA) describe the degree of relationship between the effect (main effect, interaction, linear contrast) and the dependent variable. The four most common measures of effect for the ANOVA are (1) $\eta^2$, (2) partial $\eta^2$, (3) $\omega^2$, and (4) $\rho_I$ (eta-squared, partial eta-squared, omega-squared, and intraclass correlation respectively). Eta-squared and partial eta-square primarily pertain to the relationship observed in the sample. Omega-squared and intraclass correlation estimate the degree of relationship in the population. While formulas for each of the aforementioned measures of effect will be given, the effect size measure of primary interest for this study is eta-squared.

Eta-squared describes the proportion of variance that is attributed to the effect. It can be calculated using the following formula:

(1.49) $$\eta^2 = SS_{effect}/SS_{total}$$

An issue with using eta-squared is its value is dependent on the other effects being examined. It is influenced by the other effects' magnitudes. As a solution, partial eta-squared was developed. The formula for partial eta-squared is as follows:

(1.50) $$\text{partial } \eta^2 = SS_{effect} / (SS_{effect} + SS_{error})$$

It is important to remember that though partial eta-squares are calculated for each of the effects, they are not additive by nature. They do not sum together to the amount of variance accounted for in the dependent variable by the independent variable.

Omega-squared is an estimate of the population relationship. Because it is based on the population parameters, it will be smaller then the eta-squared values (which are based on sample statistics). The formula for omega-squared is:

(1.51) $$\omega^2 = (SS_{effect} - (df_{effect})(MS_{error})) / (MS_{error} + SS_{total})$$

The final type of effect size measure used with the ANOVA is the intraclass correlation. It estimates the relationship between the independent and dependent variables in a random effects model. This effect size measure is not common in psychology. The formula for the intraclass correlation is:

(1.52) $$\rho_I = (MS_{effect} - MS_{error}) / (MS_{effect} + (df_{effect})( MS_{error}))$$

<u>Other Measures of Effect Size</u>

Currently, there are 61 different effect size measures available (Elmore, 2001). Among the 61 are measures for log-linear models / logistic regression (the odds ratio), meta-analysis (Glass's delta), nonparametric statistics, structural equation modeling, etc. These effect size measures have not been as researched as the measures mentioned above. Researchers have not formally operationalized definitions for small, medium, and large effects for these indices.

Why Report Statistical Power and Effect Size Measures

The APA has been strongly urging researchers to calculate and report both statistical power and effect size measures for years (Wilkinson, 1999). They have urged most recently to "provide information...replace calculated power in describing results" (p. 596) and to "always provide some effect size estimate when reporting a p value" (p. 599).

Not only does the APA find that these estimates are important, but other researchers have been reporting similar attitudes for years. Glass and Stanley (1970) stated, "in testing any statistical hypothesis is true or that it is false is never made with certainty; he always runs a risk of making an incorrect decision" (p. 275). Glass touches on the importance of knowledge before the research has started how likely it is that significance will be found. Sherron (1988) hypothesizes that "many 'nonsignificant' findings are the result of inadequate research design and data analysis" (p. 170). He emphasizes that calculating power can prevent running studies which have little to no chance in finding significant results.

It is not only important to calculate statistical power a priori and effect size measures upon completion, but it is paramount to psychology improving its research and research techniques. For close to a hundred years, psychologists have relied on Null Hypothesis Significance Testing

(NHST) to support their claims. This process is flawed as noted by Cohen (1969). "This is the usual expectation of the investigator, who has stated the null hypothesis for tactical purposes so that he may reject it and conclude that the phenomenon exists. But, of course, the fact that the phenomenon exists in the population far from guarantees a statistically significant result" (p.3-4). Cohen accentuated the point that just because an effect exists in nature, does not mean that the researcher has enough statistical power to detect that effect with his or her research design.

Later, Cohen (1994) explicated that though NHST is inherently flawed, there is no "magical alternative" to NHST. He concurred with the recent recommendation of the APA that psychology research must understand and improve their data (by calculating power analysis a prior) and report effect sizes (using confidence interval).

In short, there are many reasons for researchers to calculate statistical power and effect size measures. First, it adds to the body of knowledge with sound research designs. Secondly, it allows for future research to get better estimates of costs for research designs. Thirdly, it advances psychology as a science.

References

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

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

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

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

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

Glass, G., & Stanley, J. (1970). Statistical methods in education and psychology. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). What if there were no significance tests? Mahwah, NJ: Erlbaum.

Hedges, L. V., & Olkin, I. (1985). Statistical models for meta-analysis. New York: Academic Press.

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. Psychological Methods, 6, 203-217.

Hintze, J. L. (2000). PASS 2000 user's guide. Kaysville, Utah: Number Cruncher Statistical Systems.

Hunter, J. E. (1997). Needed: A ban on the significance test. Psychological Science, 8, 3-7.

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

Kosciulek, J. F., & Szymanski, E. M. (1993). Statistical power analysis of rehabilitation counseling research. Rehabilitation Counseling Bulletin, 36, 212-219.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1, 130-149.

Morrison, D. E., & Henkle, R. E. (Eds.). (1970). The significance test controversy. Chicago: Aldine.

Patnaik, P. B. (1949). The non-central $\chi^2$ and F distributions and their applications. Biometrika, 36, 203-232.

Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. Psychological Methods, 1, 331-340.

Rossi, J. S. (1984). Statistical power of psychological research. Unpublished doctoral dissertation, University of Rhode Island.

Rossi, J. S. (1988). ONEWAY: A BASIC program for computing ANOVA from group summary statistics. Behavior Research Methods, Instruments & Computers, 20, 347-348.

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? Journal of Consulting and Clinical Psychology, 58, 646-656.

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

Sherron, R. H. (1988). Power analysis: The other half of the coin. Community/ Junior College Quarterly, 12, 169-175.

Tang, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. Statistical Research Memoirs, 2, 126-149.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25, 26-30.

Thompson, B. (1997). Editorial policies regarding statistical significance tests: Further comments. Educational Researcher, 26, 29-32.

Thompson, B. (1999). Why "encouraging" effect size reporting is not working: The etiology of researcher resistance to changing practices. The Journal of Psychology, 133, 133-140.

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

Table 1.1

Possible Statistical Outcomes

| The True Nature of the Null Hypothesis in Reality | | Your Research Decision: Accept the Null Hypothesis | Your Research Decision: Reject the Null Hypothesis |
|---|---|---|---|
| | The Null Hypothesis is true in nature | Correct Decision: You would state that there isn't enough evidence to reject the Null. | Incorrect Decision: You rejected a true Null. This is a Type I error ($\alpha$). |
| | The Null Hypothesis is false in nature | Incorrect Decision: You failed to reject a false Null. This is a Type II error ($\beta$) | Correct Decision: You rejected the Null when it was false. |

Table 1.2

Definitions of Small, Medium, and Large

| | Effect Size Index | Small | Medium | Large |
|---|---|---|---|---|
| Cohen (1962) | | | | |
| *t*-test | *d*= | .25 | .50 | 1.00 |
| Pearson's *r* | *r*= | .20 | .40 | .60 |
| Cohen (1988) | | | | |
| *t*-test | *d*= | .20 | .50 | .80 |
| Pearson's *r* | *r*= | .10 | .30 | .50 |
| Diff. Betw/ Correlation | *q*= | .10 | .30 | .50 |
| Sign test | *g*= | .05 | .15 | .25 |
| Diff. Betw/ Proportions | *h*= | .20 | .50 | .80 |
| Chi-Square | *w*= | .10 | .30 | .50 |
| ANOVA | *f*= | .10 | .25 | .40 |
| ANCOVA | *f*= | .10 | .25 | .40 |
| Multiple Regression | $f^2$= | .02 | .15 | .35 |
| MANOVA | $f^2$= | .02 | .15 | .35 |
| MANCOVA | $f^2$= | .02 | .15 | .35 |

Table 1.3

Interpreting Cohen's d in terms of Percent of Nonoverlap

| Cohen's Standard | Effect Size | Percent of Nonoverlap |
|---|---|---|
| | 0 | 0 |
| | 0.1 | 7.7 |
| Small | 0.2 | 14.7 |
| | 0.3 | 21.3 |
| | 0.4 | 27.4 |
| Medium | 0.5 | 33.0 |
| | 0.6 | 38.2 |
| | 0.7 | 43.0 |
| Large | 0.8 | 47.4 |
| | 0.9 | 51.6 |
| | 1.0 | 55.4 |
| | 1.1 | 58.9 |
| | 1.2 | 62.2 |
| | 1.3 | 65.3 |
| | 1.4 | 68.1 |
| | 1.5 | 70.7 |

Table 1.4

Comparing Cohen's d, correlation, and percent of Variance Accounted for

| Cohen's Standard | Effect Size | *r* | Percent of Variance Accounted for |
|---|---|---|---|
| | 0 | 0 | 0 |
| | 0.1 | .050 | .002 |
| Small | 0.2 | .100 | .010 |
| | 0.3 | .148 | .022 |
| | 0.4 | .196 | .038 |
| Medium | 0.5 | .243 | .059 |
| | 0.6 | .287 | .083 |
| | 0.7 | .330 | .109 |
| Large | 0.8 | .371 | .138 |
| | 0.9 | .410 | .168 |
| | 1.0 | .447 | .200 |
| | 1.1 | .482 | .232 |
| | 1.2 | .514 | .265 |
| | 1.3 | .545 | .297 |
| | 1.4 | .573 | .329 |
| | 1.5 | .600 | .360 |

*All glory comes from daring to begin.*
*Eugene F. Ware*

# Chapter 2:
# A History of Statistical Power in Psychology

In the beginning of the history of statistical power, its predecessor Null Hypothesis Significance Testing (NHST) was being developed. While the idea of systematically examining data is a couple hundred years old, NHST is less than 100 years old. Some of the components of the process are older than that (the .05 level, Cowles & Davis, 1982), but the bulk of the NHST process was published and established in the 1930's.

Fisher and the team of Neyman and Pearson each separately developed a method for examining data. Fisher (1932) published the process of hypothesis testing while Neyman and Pearson (1928a, 1928b, 1933a, 1993b) published the process of significance testing. Fisher posited that a single hypothesis was necessary, while Neyman and Pearson provided for the null hypothesis and the alternative hypothesis. Fisher discussed the p-values while Neyman and Pearson provided critical values. Fisher's theory examined the data given the hypothesis ($p(\text{Data} \mid H_0)$ while Neyman and Pearson discussed fixed value probabilities. Both views are viable. Fisher and Neyman and Pearson openly loathed each other and each other's theory testing process.

Even with the emotions between the two camps, introduction to statistic textbook writers determined that a hybrid of the two methods would be best. The hybrid of the two models mentioned above is what was presented in the textbooks in the 30's and presented in the textbooks currently.

The NHST Debate

The hybrid model has been not accepted without criticism (Berkson, 1938; Hogben, 1957). Some researchers report that NHST has many flaws and others have misused it. Huberty (1993) simply proclaimed that NHST is not at fault. In his opinion, it is the faulty textbooks, the teachers and teaching system, and the editors who are to blame. Chow (1996) found that if one meets all of the assumptions of NHST then there is no problem using it. Knapp's (1998) concern with NHST is that the null hypothesis is uninformative without means and effect sizes. In general, those who support the use of NHST state that the method is not flawed – the researchers who utilize the methods are.

As mentioned above, the NHST dispute is not a new debate. It has been examined from the 30's (Berkson, 1938) to today. Books have been written to examine both sides of the issue (Harlow et al., 1997; Morrison & Henkel, 1970) and articles have been published in a variety of journals. Morrison and Henkel (1970) gathered articles from a variety of researchers (both in psychology and sociology) to present sides of the NHST argument. Researchers like Meehl had articles in that book that condemned the use of NHST (and basically anyone who used the technique). The Harlow et al. (1997) book also provides articles written by a variety of authors (Cohen, Rossi, Meehl, Rosenthal), who support and refute NHST. Some of the critics of NHST have very basic issues. Meehl (1978) essentially states that the Null hypothesis will always be false. He believes that there is no reason to examine it if it is always going to be false. Other critics mention the mass misuse of the method (Cohen, 1994). In Cohen's article "The Earth is round p< .05," he touches on a couple of issues with the misuse of the NHST method. The first issue is with the misinterpretation of the p-level. The second issue is with misinterpretation of the complement of the p-value being the probability of successful replications. Cohen also touches on the misinterpretation of null results indicating affirmation of the null hypothesis. Cohen develops the argument against null hypothesis testing by mentioning that $p(data \mid H_0) \neq p(H_0 \mid data)$. In words, the probability of the data given the null hypothesis does not equal the probability of the null hypothesis given the data.

<u>When Did Statistical Power Analysis Begin?</u>

With their controversial technique of examining data, Neyman and Pearson (1933b) attempted to separate their method out from Fisher's. In their paper presented to the Cambridge Philosophical Society, they spoke of the factor of central importance – the power of the statistical test. It was the first time that the term power was introduced. They defined two types of error in hypothesis testing. The first type of error is the Type I error. It is the error that occurs when the null is rejected (therefore concluding that there is an effect) when the null is true (because there really isn't an effect). The second type of error is the Type II error, which is when the researcher fails to reject the null when it is in fact false (there really is a treatment effect). In this same paper, Neyman and Pearson define power as the probability of rejecting the null hypothesis when it is in fact false.

The Early Research on Power

Five years after Neyman and Pearson established the term power, researchers began publishing articles which conceptualize power in terms of the statistical tests available. Tang (1938) examined the power function for the analysis of variance test. He provided some of the first tables for calculating power. Ferris, Grubbs, & Weaver (1946) also developed the first aspects of power in their article on the operating characteristics for the common statistical tests of significance. In their article, they provide some of the first power curves. Wolfowitz (1949) also provided a general article on power and tests reliant on the normal distribution. Hoeffding (1952) investigated the power of many nonparametric statistics. Stuart (1952) focused in on the power of the two difference sign test and Lehmann (1953) concentrated on the power of rank tests.

Mosteller and Bush (1954) wrote a chapter in the Handbook of Social Psychology entitled "Selected Quantitative Techniques." While the title might suggest a basic statistics overview, the chapter develops the concepts crucial in statistical power. They discuss the power of basic tests such as the

F-test, t-test, and chi-square. A couple years after, Linhart (1957) operationalized the power functions of tests concerning the product moment correlation coefficient.

Introducing Power to Psychology

The researcher who has been reported as being responsible for introducing statistical power to modern psychology is Jacob Cohen. Cohen (1962) conducted the first statistical power survey of psychological research. He examined two popular journals and analyzed the results of the studies to determine post hoc power. From his analysis, Cohen determined that most of the research reported in the journals he examined did not have sufficient power to achieve statistically significant results.

Cohen proceeded to make statistical power accessible to psychologists and researchers alike. In his groundbreaking textbook on the issue (most recent edition 1988), he provided statistical tables for statistics such as the t-test, correlation, differences between correlation coefficients, chi-square, ANOVA, and some multivariate techniques. Not only did Cohen expound on the basics of statistical power, but he also provided tables for sample size calculations so that researchers could develop studies with sufficient power. As if the text was not enough, Cohen continued to write on the subject and provided a statistical "primer" on power in 1992.

Following in Cohen's Footsteps

After Cohen's seminal work, researchers in the 1970's began to analyze other research journals using the power survey method that Cohen pioneered. Ten years after Cohen, Brewer (1972) examined the American Educational Research Journal. The results were similar to Cohen's, in that most articles did not have sufficient power to detect small effects sizes, had about a 50-50 chance of detecting medium effects sizes, and on average sufficient power to identify large effect sizes. Additional power surveys found comparable results. Power surveys were done on Research Quarterly (Jones and Brewer, 1972), the Journal of Research in Science Teaching (Pennick & Brewer, 1972), the Journal of Educational Measurement (Brewer & Owen, 1973), the Journal of Communication (Kattzer and Sodt, 1973), Counselor Education and Supervision (Haase, 1974), American Sociological Review (Spreitzer & Chase, 1974), the American Forensic Association Journal (Chase & Tucker, 1975), the Journal of Communication Disorders (Kroll & Chase, 1975), the Journal of Speech and Hearing Research (Kroll & Chase, 1975), Journalism Quarterly (Chase & Baran, 1976), Journal of Broadcasting (Chase & Baran, 1976), the Journal of Applied Psychology (Chase & Chase, 1976), Research Quarterly (Christensen & Christensen, 1977), and the American Journal of Occupational Therapy (Chase, Chase, & Tucker, 1978). After the power surveys of the '70's, a couple power surveys were conducted in various areas in the '80s and '90s. Though one might assume with the advent of the personal computer and computer programs for conducting power analysis that power surveys might increase in popularity, the appeal of power surveys have seemingly died off. A comprehensive listing of power surveys can be found in table 2.1.

---

Insert Table 2.1 here

---

To determine if Cohen's work on power analysis had influenced how researchers did their research, Joseph Rossi preformed a follow up power survey on similar journals to Cohen's. Rossi (1990) found that statistical power had not increased since Cohen's original study. It seems that researchers in the eighties still did not calculate a priori the probability of achieving a significant result.

While power surveys are fewer in the literature, they are becoming more specified in nature. A recent power survey (Kazantzis, 2000) examined the statistical power of psychotherapy outcome research. In contrast to the earlier power surveys, Kazantzis did not focus his power survey on one specific or even a couple specified journals. His focus was on the topic area and determining if the research area was built on sound research.

Advancements in Power Analysis

As mentioned earlier, the advent of the personal computer has made statistical power even more accessible. With relative ease, a researcher can determine the correct sample size a priori for a specified alpha level, based on the statistic, and expected effect using programs such as GPOWER (Buchner, Erdfelder, & Faul, 1997), PASS (NCSS, 2001), or Rossi's BASIC programs (1984, 1988).

Researchers have also expounded on Cohen's earlier research by developing statistical techniques for calculating power for a variety of statistical methods. For example, researchers have developed methods for determining the power of structural equation models (MacCallum, Browne, & Sugawara, 1996), the power of randomization tests with multiple baseline designs (Ferron & Sentovich, 2002), repeated measures designs (D'Amico, Neilands, & Zambarano, 2001), meta-analysis (Hedges & Pigott, 2001), Log-Linear Modeling (Schuster & von Eye, 2000), split-plot designs (Bradley & Russell, 1998), dichotomous moderator variables (Aguinis, Pierce, & Stone-Romero, 1994), and configural frequency analysis (Indurkhya & von Eye, 2000).

While research on the concepts of statistical power are continuing, it seems that power surveys and power research are not having an effect on psychological research (Sedlmeier & Gigerenzer, 1989). Even in the most recent power surveys (Kenna & Rossi, 2002), the research articles examined did not have enough power to identify a medium effect size.

## The Future of Statistical Power Analysis

Presently, the APA (2001) strongly encourages a priori statistical power calculations in the most recent version of the publication manual. They also suggest the reporting of power in research articles. Grant applications also require estimates of statistical power. It would seem that with the recommendations and requirements of power calculations that the statistical power in current research would be adequate to detect significant effects. It would also seem that with the push for the calculation of statistical power that students of psychology would be exposed to the concept of power and possibly simple power calculations. Given the recent endorsement of statistical power, one might assume that in the near future that journal might require the reporting of statistical power and that statistical power would be a common topic in statistical textbooks.

References

Acklin, M. W., McDowell, C. J., & Orndoff, S. (1992). Statistical power and the Rorschach: 1975-1991. Journal of Personality Assessment, 59, 366-379.

Aguinis, H., Pierce, C. A., & Stone-Romero, E. F. (1994). Estimating the power to detect dichotomous moderators with moderated multiple regression. Educational & Psychological Measurement, 54, 690-692.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the Chi-square test. Journal of the American Statistical Association, 33, 526-536.

Bezeau, S. & Graves, R. (2001). Statistical power and effect sizes of clinical neuropsychology research. Journal of Clinical & Experimental Neuropsychology, 23, 399-406.

Brewer, J. K. (1972). On the power of statistical tests in the American Educational Research Journal. American Educational Research Journal, 9, 391-401.

Brewer, J. K., & Owen, P. W. (1973). A note on the power of statistical tests in the Journal of Educational Measurement. Journal of Educational Measurement, 10, 71-74.

Brown, J. & Hale, M. S. (1992). The power of statistical studies in consultation-liason psychiatry. Psychosomatics, 33, 437-443.

Buchner, A., Erdfelder, E., & Faul, F. (1997). How to Use G*Power. Retrieved on April 23, 2002 from http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/how_to_use_gpower.html.

Cady-Webster, K., Hevey, C., Huang, J., & Rossi, J. (2000). Not all power corrupts: Statistical power of research published in the psychology of women quarterly. Paper presented at the Annual Convention of the Eastern Psychological Association, Baltimore, MD.

Chase, L. J., & Chase, R. B. (1976). A statistical power analysis of applied psychological research. Journal of Applied Psychology, 61, 234-237.

Chase, L. J., & Tucker, R. K. (1975). A power-analytic examination of contemporary communication research. Speech Monographs, 42, 29-41.

Chase, L. J., & Tucker, R. K. (1976). Statistical power: Derivation, development, and data-analytic implications. Psychological Record, 26, 473-486.

Chow, S. L. (1998). Precis of statistical significance: Rationale, validity, and utility. Behavioral and Brain Sciences, 21, 169-239.

Christensen, J. E., & Christensen, C. E. (1977). Statistical power analysis of health, physical education, and recreation research. Research Quarterly, 48, 204-208.

Clark-Carter, D. (1997). The account taken of statistical power in research published in the British Journal of Psychology. British Journal of Psychology, 88, 71-83.

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

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

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

Cowles, M. P., & Davis, C. (1982). On the origins of the .05 level of statistical significance. American Psychologist, 37, 553-558.

D'Amico, E. J., Neilands, T. B., & Zambarano, R. (2001). Power analysis for multivariate and repeated measures designs: A flexible approach using the SPSS MANOVA procedure. Behavior Research Methods, Instruments & Computers, 33, 479-484.

Daly, J. A., & Hexamer, A. (1983). Statistical power in research in English education. Research in the Teaching of English, 17, 157-164.

Dilullo, L. K. (1998). A post hoc power analysis of inferential research examining the relationship between mathematics anxiety and mathematics performance. (Doctoral Dissertation, Auburn University, 1998). Dissertation Abstracts International, 58, 2574.

Ferris, C. D., Grubbs, F. E., & Weaver, C. L. (1946). Operating characteristics for the common statistical tests of significance. Annals of Mathematical Statistician, 17, 178-197.

Ferron, J., & Sentovich, C. (2002). Statistical power of randomization tests used with multiple-baseline designs. Journal of Experimental Education, 70, 165-178.

Fisher, R. A. (1932). Statistical methods for research workers (4th ed.). London: Oliver & Boyd.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). What if there were no significance tests? Mahwah, NJ: Erlbaum.

Haase, R. F. (1976). Power analysis of research in counselor education. Counselor Education and Supervision, 14, 124-132.

Hedges, L.V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. Psychological Methods, 6, 203-217.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. Annals of Mathematical Statistics, 23, 169-192.

Hogben, L. (1957). Statistical theory: The relationship of probability, credibility, and error. An examination of the contemporary crisis in statistical theory from a behaviorist viewpoint. London: Allen & Unwin.

Indurkhya, A., & von Eye, A. (2000). The power of tests in Configural Frequency Analysis. Psychologische Beitrage, 42, 301-308.

Jones, B. J., & Brewer, J. K. (1972). An analysis of the power of statistical tests reported in the Research Quarterly. Research Quarterly, 43, 23-30.

Katzer, J., & Sodt, J. (1973). An analysis of the use of statistical testing in communication research. Journal of Communication, 23, 251-265.

Kazantzis, N. (2000). Power to detect homework effects in psychotherapy outcome research. Journal of Consulting & Clinical Psychology, 68, 166-170.

Kazdin, A. E., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. Journal of Consulting & Clinical Psychology, 57, 138-147.

Kenna, G., & Rossi, J. S. (2002, March) Statistical power in the journal "Experimental and Clinical Psychopharmacology." Poster session presented at the 73rd Eastern Psychological Association convention, Boston, MA.

Knapp, T. R. (1998) Comments on the statistical significance testing articles. Research in the Schools, 5, 39-41.

Kosciulek, J. F. (1993). The statistical power of vocational evaluation research. Vocational Evaluation and Work Adjustment Bulletin, 26, 142-145.

Kosciulek, J. F., & Szymanski, E. M. (1993). Statistical power analysis of rehabilitation counseling research. Rehabilitation Counseling Bulletin, 36, 212-219.

Kroll, R. M., & Chase, L. J. (1975). Communication disorders: A power-analytic assessment of recent research. Journal of Communication Disorders, 8, 237-247.

Lehmann, E. L. (1953). The power of rank tests. Annals of Mathematical Statistics, 24, 23-43.

Levenson, R. L. (1980). Statistical power analysis: Implications for researchers, planners, and practitioners in gerontology. Gerontologist, 20, 494-498.

Linhart, H. (1957). Power functions of tests concerning the product moment correlation coefficient. Journal National Institutional Personnel Research, 7, 51-65.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1, 130-149.

Maddock, J. E., & Rossi, J. S. (2001). Statistical power of articles published in three health-psychology related journals. Health Psychology, 20, 76-78.

Mazen, A. M., Graf, L. A., Kellog, C. E., & Hemmasi, M. (1987). Statistical power in contemporary management research. Academy of Management Journal, 30, 1105-1112.

Mazen, A. M., Hemmasi, M., & Lewis, M. F., (1987). Assessment of statistical power in contemporary strategy research. Strategic Management Journal, 8, 403-410.

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

Mone, M. A., Mueller, G. C., & Mauland, W. (1996). The perceptions and usage of statistical power in applied psychology and management research. Personnel Psychology, 49, 103-120.

Morrison, D. E., & Henkle, R. E. (Eds.). (1970). The significance test controversy. Chicago: Aldine.

Mosteller, F. & Bush, R. R. (1954). Selected quantitative techniques. In G. Lindzey (Ed.), Handbook of Social Psychology: Theory and Method (Vol. 1, pp. 289-334). Reading, MA: Addison-Wesley.

Neyman, J., & Pearson, E. S. (1928a). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika, 20A, 175-240.

Neyman, J., & Pearson, E. S. (1928b). On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. Biometrika, 20A, 263-294.

Neyman, J., & Pearson, E. S. (1933a). On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London, Series A, 231, 289-337.

Neyman, J., & Pearson, E. S. (1933b). The testing of statistical hypotheses in relation to probabilities a priori. Proceedings of the Cambridge Philosophical Society, 29, 492-510.

Orme, J. G., & Combs-Orme, T. D. (1986). Statistical power and Type II errors in social work research. Social Work Research & Abstracts, 22, 3-10.

Orme, J. G., & Tolman, R. M. (1986). The statistical power of a decade of social work education research. Social Service Review, 60, 620-632.

Ottenbacher, K. (1982). Statistical power of research in occupational therapy. Occupational Therapy Journal of Research, 2, 13-25.

Pennick, J. E., & Brewer, J. K. (1972). The power of statistical tests in science teaching research. Journal of Research in Science Teaching, 9, 377-381.

Reed, J. F., & Slaichert, W. (1980). Statistical proof in inconclusive 'negative' trials. Archives of Internal Medicine, 141, 1307-1310.

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? Journal of Consulting and Clinical Psychology, 58, 646-656.

Sanchez, J., Valera, A., Velandrino, A. P. & Marin, F. (1992). Un studio de la potencia estadistica en Anales de Psicologia (1984 - 1991). Anales de Psicologia, 8, 19-32.

Sawyer, A. G., & Ball, A. D. (1981). Statistical power and effect size in marketing research. Marketing Research, 18, 275-290.

Schuster, C., & von Eye, A. (2000). Using Log-Linear Modeling to increase power in two-sample Configural Frequency Analysis. Psychologische Beitrage, 42, 273-284.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? Psychological Bulletin, 105, 309-316.

Sindelar, P.T., Allman, C. Monda, L., Vail, C. O., Wilson, C. L. & Schloss, P. J. (1988). The power of hypothesis testing in special education efficacy research. Journal of Special Education, 22, 284-296.

Spreitzer, E., & Chase, L. J. (1974). Statistical power in sociological research: An examination of data-analytic strategies. Unpublished manuscript, Bowling Green State University, Department of Sociology, Bowling Green, Oh.

Stuart, A. (1952). The power of two difference-sign tests. Journal of the American Statistical Association, 47, 416-424.

Tang, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. Statistical Research Memoirs, 2, 126-149.

Valera, A., Sanchez, J., Marin, F., & Velandrino, A. P. (1998). Potencia estadistica de la Revista de Psicologia General y Aplicada. Revista de Psicologia General y Aplicada, 51, 233-246.

Whittington, C. J., Podd, J., & Kan, M. M. (2000). Recognition memory impairment in Parkinson's disease: Power and meta-analyses. Neuropsychology, 14, 233-246.

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

Wolfowitz, J. (1949). The power of the classical tests associated with the normal distribution. Annals of Mathematical Statistics, 20, 540-551.

Wooley, T.W. (1983). A comprehensive power-analytic investigation of research in medical education. Journal of Medical Education, 58, 710-715.

Wooley, T. W., & Dawson, G. O. (1983). A follow-up power analysis of the tests used in Journal of Research in Science Teaching. Journal of Research in Science Teaching, 20, 673-681.

Table 2.1

Statistical Power Research Surveys

| Source | Journals included in the surveys | Years covered | Sample Size | | Statistical Power Estimates | | |
|---|---|---|---|---|---|---|---|
| | | | Articles | Tests | Small Effects | Medium Effects | Large Effects |
| Cohen (1962) | Journal of Abnormal and Social Psychology | 1960 | 70 | 2,088 | .18 | .48 | .83 |
| Brewer (1972) | American Educational Research Journal | 1969-1971 | 47 | 373 | .14 | .58 | .78 |
| Jones & Brewer (1972) | Research Quarterly | 1969-1971 | 106 | 261 | .15 | .54 | .83 |
| Pennick & Brewer (1972) | Journal of Research in Science Teaching | 1969-1970 | 66 | 554 | .22 | .71 | .87 |
| Brewer & Owen (1973) | Journal of Educational Measurement | 1969-1971 | 13 | 267 | .21 | .72 | .96 |
| Katzer & Sodt (1973) | Journal of Communication | 1971-1972 | 31 | 1,671 | .23 | .56 | .79 |
| Haase (1974) | Counselor Education and Supervision | 1968-1971 | 60 | 206 | .19 | .46 | .72 |
| Spreitzer & Chase (1974) | American Sociological Review | 1972-1973 | 34 | 1,049 | .55 | .84 | .94 |
| Chase & Tucker (1975) | American Forensic Association Journal | 1973 | 46 | 1,298 | .18 | .52 | .79 |
| | Central States Speech Journal | 1973 | | | | | |
| | Journal of Communication | 1973 | | | | | |
| | Quarterly Journal of Speech | 1973 | | | | | |
| | Southern Speech Journal | 1973 | | | | | |
| | Speech Monographs | 1973 | | | | | |
| | Speech Teacher | 1973 | | | | | |
| | Today's Speech | 1973 | | | | | |
| | Western Speech | 1973 | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Kroll & Chase (1975) | Journal of Communication Disorders | 1973-1974 | 62 | 1,037 | .16 | .44 | .73 |
| | Journal of Speech and Hearing Research | 1973-1974 | | | | | |
| Chase & Baran (1976) | Journalism Quarterly | 1974 | 48 | 701 | .34 | .76 | .91 |
| | Journal of Broadcasting | 1974 | | | | | |
| Chase & Chase (1976) | Journal of Applied Psychology | 1974 | 121 | 3,373 | .25 | .67 | .86 |
| Christensen & Christensen (1977) | Research Quarterly | 1975 | 43 | NR | .18 | .39 | .62 |
| Chase, Chase, & Tucker (1978) | American Journal of Occupational Therapy | 1980 | 25 | 3,304 | .38 | .62 | .81 |
| Levenson (1980) | Gerontologist | 1961-1977 | 56 | NR | .37 | .88 | .96 |
| | Journal of Gerontology | 1946-1977 | | | | | |
| Reed & Slaichert (1981) | American Journal of Surgery | 1977 | 355 | 2,619 | .138 | .387 | .614 |
| | American Journal of Medicine | 1977 | | | | | |
| | New England Journal of Medicine | 1977 | | | | | |
| | American Journal of Cardiology | 1977 | | | | | |
| | Journal of Pediatrics | 1977 | | | | | |
| | American Review of Respiratory Disease | 1977 | | | | | |
| Sawyer & Ball (1981) | Journal of Marketing Research | 1979 | 23 | 475 | .41 | .89 | .98 |
| Ottenbacher (1982) | American Journal of Occupational Therapy | 1980 | 22 | 205 | .37 | .65 | .93 |
| Daly & Hexamer (1983) | Research in the Teaching of English | 1978-1980 | 57 | 1,233 | .22 | .63 | .86 |
| Woolley & Dawson (1983) | Journal of Research in Science Teaching | 1977-1980 | 192 | 3,556 | .23 | .63 | .85 |
| Woolley (1983) | Journal of Medical Education | 1980-1982 | 100 | 2,220 | .23 | .69 | .90 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Orme & Combs-Orme (1986) | Social Work Research Abstracts | 1977-1984 | 49 | 3,114 | .31 | 76 | .92 |
| Orme & Tolman (1986) | Journal of Social Work Education | 1976-1985 | 64 | 1,998 | .20 | .68 | .88 |
| Mazen, Hemmasi, & Lewis (1987) | Strategic Management Journal | 1982-1984 | 44 | 3,665 | .23 | .59 | .83 |
| | Academy of Management Journal | 1984 | | | | | |
| Mazen, Graf, Kellogg, & Hemmasi (1987) | Academy of Management Journal | 1984 | 84 | 7,215 | .31 | .77 | .91 |
| | Journal of Management | 1984 | | | | | |
| | Proceedings of the Midwest Division of the Academy of Management | 1984 | | | | | |
| Sindelar, Allman, Monda, Vail, Wilson, & Schloss (1988) | Special Education Efficacy Research | NR | 44 | 26 | .12 | .46 | .79 |
| | | | | 27 | .12 | .45 | .76 |
| Kazdin & Bass (1989) | Comparative Psychotherapy Outcome Research | 1984-1986 | 85 | 2,501 | NR | NR | NR |
| Sedlmeier & Gigerenzer (1989) | Journal of Abnormal Psychology | 1984 | 54 | NR | .21 | .50 | .84 |
| McKean (1990) | Educational Psychology Ph.D. Dissertations | 1988 | NA | NA | .17 | .54 | .80 |
| Rossi (1990) | Journal of Abnormal Psychology | 1982 | 49 | 1,289 | .16 | .56 | .84 |
| | Journal of Consulting and Clinical Psychology | 1982 | 78 | 2,231 | .18 | .58 | .83 |
| | Journal of Personality and Social Psychology | 1982 | 94 | 2,635 | .16 | .55 | .81 |
| Acklin, McDowell, & | Journal of Personality Assessment | 1975-1991 | 158 | NR | .13 | .56 | .85 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Orndoff (1992) | | | | | | | |
| | Journal of Consulting and Clinical Psychology | 1975-1991 | | | | | |
| | Journal of Clinical Psychology | 1975-1991 | | | | | |
| | Journal of Abnormal Psychology | 1975-1991 | | | | | |
| | Psychological Bulletin | 1975-1991 | | | | | |
| | American Journal of Psychiatry | 1975-1991 | | | | | |
| | Journal of Personality and Social Psychology | 1975-1991 | | | | | |
| Brown & Hale (1992) | Psychosomatics | 1989 | 24 | NR | .19 | .60 | .84 |
| Sanchez, Valera, Velandrino, & Marin (1992) | Anales de Psicologia | 1984-1991 | 16 | NR | .13 | .47 | .76 |
| Kosciulek (1993) | Vocational Evaluation and Work Adjustment Bulletin | 1989-1991 | 14 | NR | .14 | .56 | .84 |
| Kosciulek & Szymanski (1993) | Rehabilitation Counseling Bulletin | 1990-1991 | 32 | NR | .15 | .63 | .90 |
| | Rehabilitation Psychology | 1990 | | | | | |
| | Journal of Rehabilitation | 1990 | | | | | |
| | Journal of Applied Rehabilitation Counseling | 1990 | | | | | |
| | Rehabilitation Education | 1990 | | | | | |
| Mone, Meuller, & Mauland (1996) | Journal of Applied Psychology | 1992-94 | 30 | 100 | .35 | .82 | .95 |
| | Personnel Psychology | 1992-94 | 30 | 105 | .30 | .83 | .97 |
| | Organizational Behavior and Human Decision Process | 1992-94 | 30 | 113 | .17 | .60 | .87 |
| Clark-Carter (1997) | British Journal of Psychology | 1993-1994 | 54 | 1,090 | .20 | .60 | .82 |
| Dilullo (1998) | Journal for Research in Mathematics | 1976-1995 | NA | 81 | | .81 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Education | | | | | | |
| Valera, Sanchez, Marin, & Velandrino (1998) | Revista de Psicologia General y Aplicada | 1990-1992 | 89 | NR | .17 | .57 | .83 |
| Cady-Webster, Hevey, Huang, & Rossi (2000) | Psychology of Women Quarterly | 1996-1999 | 58 | 761 | .31 | .85 | .97 |
| Kazantzis (2000) | Psychotherapy Outcome Research | 1980-1998 | 27 | 32 | .11 | .44 | .71 |
| Whittington et al. (2000) | Memory Impairment in Parkinson's Disease | 1978-1997 | 46 | 1,360 | .20 | .63 | .85 |
| Bezeau & Graves (2001) | Journal of Clinical and Experimental Neuropsychology<br>Journal of International Neuropsychology Society<br>Neuropsychology | 1998-1999 | 66 | NR | NR | .451 | .785 |
| Maddock & Rossi (2001) | Journal of Studies on Alcohol | 1997 | 61 | 3,388 | .41 | .81 | .92 |
| | Health Psychology | 1997 | 56 | 2,429 | .34 | .74 | .92 |
| | Addictive Behaviors | 1997 | 70 | 2,449 | .34 | .75 | .90 |
| Kenna & Rossi (2002) | Experimental and Clinical Psychopharmacology | 1999-2000 | 48 | 619 | .12 | .36 | .59 |

*All truth passes through three stages.*
*First, it is ridiculed.*
*Second, it is violently opposed.*
*Third, it is accepted as being self-evident.*
*- Arthur Schopenhauer (1788-1860)*

# Chapter 3:

## A History of Effect Sizes

Over the years, the American Psychological Association (APA) has made a concerted effort in encouraging researchers to provide an indication of an effect size when reporting their statistical results (Hogarty & Kromrey, 2001). Effect size measures have been seen as a vital complement to tests of significance. Despite recent urgings for the consistent and methodical reporting of effect size indices, these measures are rarely found in published articles and do not appear to be standard practice (Kirk, 1996; Thompson & Snyder, 1997, 1998). Hogarty and Kromrey (2001) contend, "the reporting of effect sizes assists researchers in planning future research (i.e., the determination of sample size for subsequent experimentation) as well as facilitating comparison of results across studies through the use of meta-analytic techniques."

The current argument has influenced a number of editors to require effect size reporting. A total of 20 journals now require effect size reporting (Huberty, 2002). They are: (1) Career Development Quarterly, (2) Contemporary Educational Psychology, (3) Early Childhood Research Quarterly, (4) Educational and Psychological Measurement, (5) Exceptional Children, (6) Journal of Agricultural Education, (7) Journal of Applied Psychology, (8) Journal of Community Psychology, (9) Journal of Consulting & Clinical Psychology, (10) Journal of Counseling and Development, (11) Journal of Early Intervention, (12) Journal of Educational and Psychological Consultation, (13) Journal of Experimental Education, (14) Journal of Learning Disabilities, (15) Language Learning, (16) Measurement and Evaluation in Counseling and Development, (17) The Professional Educator, (18) Reading and Writing, (19) Research in Schools, and (20) Journal of Personality Assessment.

While the most common reporting of effect sizes is effect sizes as applied to univariate comparisons, there are effect size measures available for a broad range of statistics including, but not limited to: prediction methods, multiple regression, MANOVA, and proportion comparisons.

Correlation

One of the first indices established to examine the relationship between variables was the correlation. While its first use is under dispute (Cowles, 1989; Stigler, 1986), it was either developed by Sir Francis Galton (Hald, 1998) or by his cousin, Charles Darwin (Cowles, 1989). Whether it was developed and presented by Galton or Darwin first is difficult to determine, but it is evident that the discovery was made in the late 1800's. Soon after in 1892, Edgeworth used the symbol $\rho$ for the *coefficient of correlation*. Around 1896, Pearson used $r$ for the same concept. In 1905, Pearson developed the concept of the *correlation ratio*, $\eta$. In 1924, Fisher used Pearson's $\eta$ and derived a probability distribution for it in terms of the analysis of variance. Kelly (1935) adjusted this statistic and proposed $\varepsilon^2$.

Later, researchers made a connection between the effect size indices of $\eta^2$ and $\varepsilon^2$ and the analysis of variance (Peters & Van Voorbis, 1940).

Soon researchers sought to improve upon $\eta^2$ and reduce some of its estimation bias. After $\varepsilon^2$, Hays (1963) offered yet another solution the effect size index, est. $\omega^2$. This index estimates the strength of the relationship between the grouping variable and a dependent variable.

While correlations and other indices of relationship have been considered to examine effect size, researchers commonly square the correlation value to determine the percent of variance accounted for in the dependent variable by the independent variable. Cohen (1969) operationalized the shared variance and assigned small, medium, and large labels to $r^2$ values. Rosenthal and Rubin (1979) criticized these labels and later (1982) offered a solution. Their solution was the binomial effect size display (BESD).

If the relationship being examined contains two dichotomous variables, Yule's (1900) Q is the appropriate index. Other options are Pearson coefficient of mean square contingency, Pearson tetrachoric coefficient of correlation, and Tschuprow coefficient (Cowles, 1989). Cramer's (1946) C can be used in this situation and can also be used when comparing multiple proportions.

Group Differences

Cohen (1962) established one of the first indices to examine two group mean difference. His index is Cohen's d. Hays (1963) also contributed to this area with his $\delta$ and $\omega^2$. Over time, researchers sought to improve Cohen's d by selecting different standard deviations to be in the denominator. Cohen (1969) decided on pooled standard deviation for the denominator of his index while Glass (1976) proposed the standard deviation of the control group. Once again a researcher sought to reduce the bias in this estimator and developed a new index. Hedges (1981) developed g in this effort. Another alternative to d and g, is trimmed means or Winsorized variances (Yuen, 1974). Recently, researchers developed a "common language" (CL: McGraw & Wong, 1992) statistic that expresses the relative frequency with which a score from one distribution will be greater than a score sample for the second distribution.

Cohen (1962) also developed an index for multiple group comparisons, f. During that same year, Winer (1962) established an index for estimating the effect of the treatment, $\tau_j$. Cohen (1969) continued to develop effect size indices and actually proposed $\delta$ for a standardized mean difference when working with two or more groups and (1962) a simple difference in proportions ($|P_1 - P_2|$) when working with a dichotomous dependent variable.

Group Overlap Indices

When examining how much the groups overlap, the indices of choice stem from the work of Kelley (1920, 1923) and Tilton (1937). Tilton asserted that whenever a researcher is comparing means the results "should be supplemented whenever possible by an explicit measure of overlapping, such as the percentage of area common to the two distributions" (p. 657). After thirty years, researchers began to restate Tilton's original assertion (Alf & Abrahams, 1968; Dunnette, 1966).

Multivariable Effect Indices

Indices were also developed for research situations when multiple variables were used. Pearson and Lee (1897) introduced the concept of multiple correlation and Pearson (1914) extended this work to what he would call R (*coefficient of multiple correlation*). Cohen (1977) established $f^2$ as an effect size index. Cohen's $f^2$ is related to R in the following way:

$$f^2 = R^2 / (1 - R^2) \tag{3.1}$$

His index is considered a signal to noise ratio. Huberty (1994) sought to improve this index with an adjusted $R^2$ value.

Tatsuoko (1970) is credited with establishing the first multivariate effect size index (Huberty, 2002). The effect size index proposed for the MANOVA is:

$$\eta^2 = 1 - \Lambda \tag{3.2}$$

where $\Lambda$ refers to Wilk's (1932) MANOVA criterion. Similar to the adjustments made on the univariate versions of effect size, Tatsuoka (1973) offered corrections in estimation for the multivariate counterparts.

Effect Size Surveys

While effect size estimations have been around for over 100 years, seemingly few people have examined the magnitude of effects in research. Cohen (1962) made judgments based on his experience to what he dubbed small, medium, and large effects, but few people have surveyed the research to determine actual estimates of small, medium, and large effect sizes.

The earliest effect size survey was completed by Hamblin (1971) and appraised the magnitude of effects for the first here issues of American Sociological Review in 1961. He report average effect size (.107) in terms of explained variance (V). Brown (1975) examined the distribution of Pearson's correlation and R-squared in the American Educational Research Journal from 1970 to 1974. If one was to convert his findings to eta-squared, Brown found a mean effect size of .083. Craig, Eison, & Metze (1976) examined 62 articles from the Journal of Educational Psychology, the Journal of Comparative and Physiological Psychology, and the Journal of Personality and Social Psychology. For each article, a value of omega squared was calculated if it was not reported. Craig et al. (1976) did not report summary indices of fit for the articles that they surveyed. Cooper and Findley (1982) calculated for effect size indices (d, f, r, and w) for Social Psychology textbooks. They found 1.19, .45, .60, .48, and .26 for d, f (df=1), f (df>1), r, and w respectively. Haase, Waechter, & Solomon (1982) examined the research in the Journal of Counseling Psychology from 1970 to 1979. They found a median of .083 for eta-squared in their research. Rubin and Conway (1985) examined 10 journals and estimated the effect sizes for the research articles. They examined journals between 1981-1983 and found a median effect size of .130. Most recently, Thompson and Snyder (1998) examined the 1996 issues of the Journal of Counseling & Development. They found a mean PVA of .148 (SD = .134).

While effect sizes have continued to progress over the last 100 years, few researchers have estimated the average or median effect size in their area of interest (figure 3.1 presents a history of effect size). Meta-analytic procedures have introduced methods for determining average effect sizes for research areas, but in terms of verifying Cohen's small, medium, and large labels, that research has yet to be done.

---

Insert Figure 3.1 here

---

## References

Alf, E., & Abrahams, N. M. (1968). Relationship between percent overlap and measure of correlation. Educational and Psychological Measurement, 28, 779-792.

Brown, D. J. (1975). Mirror, mirror... down with the linear model. American Educational Research Journal, 12, 491-505.

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

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

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

Cooper, H., & Findley, M. (1982). Expected effect sizes: Estimates for statistical power analysis in social psychology. Personality and Social Psychology Bulletin, 8, 168-173.

Cowles, M. (1989). Statistics in psychology: A historical perspective. Hillsdale, NJ: Lawrence Erlbaum.

Craig, J. R., Eison, C. L., & Metze, L. P. (1976). Significance testes and their interpretation: An example utilizing published research and $\omega^2$. Bulletin of the Psychonomic Society, 7, 280-282.

Cramer, H. (1946). Mathematical methods of statistics. Princeton, NJ: Princeton University Press.

Dunnette, M. D. (1966). Personnel selection and placement. Belmont, Ca: Wadsworth.

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

Hamblin, R. L. (1971). Mathematical experimentation and sociological theory: A critical analysis. Sociometry, 34, 423-452.

Hays, W. L. (1963). Statistics for psychologists. New York: Holt, Rinehart, & Winston.

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

Hogarty, K. Y., & Kromrey, J. D. (2001, April). We've bee reporting some effect sizes: Can we guess what they mean? Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Huberty, C. J. (1994) A note on interpreting an $R^2$ value. Journal of Educational and Behavioral Statistics, 19, 351-356.

Huberty, C. J. (2002) A history of effect size indices. Educational and Psychological Measurement, 62, 227-240.

Kelley, T. L. (1920). Measurement of overlapping. Journal of Educational Psychology, 11, 458-461.

Kelley, T. L. (1923). Statistical method. New York: Macmillan.

Kelley, T. L. (1935). An unbiased correlation ratio. Proceedings of the National Academy of Sciences, 21, 554-559.

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

McGraw, K. O. & Wong S. P. (1992). The descriptive use of absolute differences between pairs of scores with a common mean and variance. Journal of Educational Statistics, 19,103-110.

Pearson, K. (1914). On certain errors with regard to multiple correlation occasionally made by those who have not adequately studies this subject. Biometrika, 10, 181-187.

Pearson, K., & Lee, A. (1897). On the distribution of frequency (variation and correlation) of the barometric height of divers stations. Philosophical Transactions of the Royal Society of London, 190, 423-469.

Peters, C. C., & Van Voorhis, W. R. (1940). Statistical procedures and their mathematical bases. New York: Mc Graw-Hill.

Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. Journal of Applied Social Psychology, 9, 395-396.

Rubin, A., & Conway, P. G. (1985). Standards for determining the magnitude of relationships in social work research. Social Work Research & Abstracts, 21, 34-39.

Stigler, S. M. (1986). The history of statistics. Cambridge, MA: Belknap.

Tatsuoka, M. M. (1970). Discriminant analysis: The study of group differences. Champaign, IL: Institute for Personality and Ability Testing.

Tatsuoka, M. M. (1973). An examination of the statistical properties of a multivariate measure of strength of association. Final Report to U.S. Office of Education on Contract No. OEG-5-72-0027.

Thompson, B., & Snyder, P. A. (1997). Statistical significance testing practices in the Journal of Experimental Education. Journal of Experimental Education, 66, 75-83.

Thompson, B., & Snyder, P. A. (1998). Statistical significance and reliability analyses in recent JCD research articles. Journal of Counseling and Development, 76, 436-441.

Tilton, J. W. (1937). The measurement of overlapping. Journal of Eduacational Psychology, 28, 656-662.

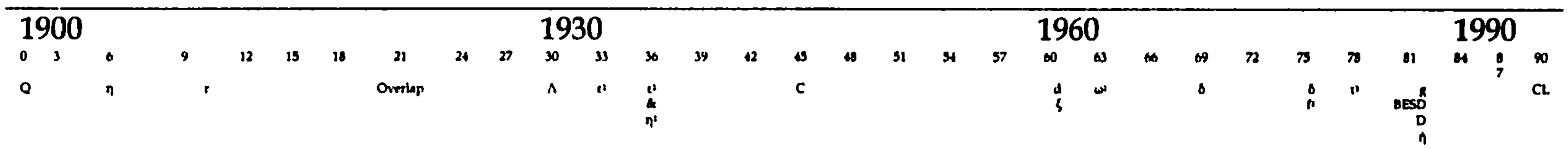
Wilks, S. S. (1932). Certain generalizations of the analysis of variance. Biometrika, 39, 471-494.

Winer, B. J. (1962). Statistical principles in experimental design. New York: McGraw-Hill.

Yule, G. U. (1900). On the association of attributes in statistics. Philosophical Transactions of the Royal Society, A, 194, 257-319.

Figure 3.1

Historical Developments in Effect Size Measures (adapted from Huberty, 2002)



*Sometimes in the winter*
*you see things that*
*you can't see in any other season.*
*Billie Jean Ward*

# Chapter 4:
# Examining the Statistical Power and Effect Sizes in Three Psychological Journals

From the beginning of psychology, researchers have been evaluating hypotheses using statistical methods. Most commonly, researchers evaluate their hypotheses against the probability of making a Type I error. Over the last 100 years, individuals have attempted to persuade researchers to augment their findings with statistical power calculations, effect sizes, and confidence intervals (Cohen, 1965; Hayes, 1963; Thompson, 1997a, 1997b, 1999, 2002). Yet, these suggestions have been for the most part ignored (Thompson, 1999).

Part of the discussion concerning evaluating hypotheses involves the dispute concerning significance testing. The debate concerning the utility of Null Hypothesis Significance Testing (NHST) is not a new one. It has been examined, criticized, and supported (Berkson, 1938; Chow, 1996; Harlow et al., 1997; Hogben, 1957; Morrison & Henkel, 1970) since the first psychology textbooks presented the hybrid method. Some researchers insist that tests of statistical significance are not useful (Carver, 1978, 1993; Cohen, 1994; Hunter, 1997; Kirk, 1996; Schmidt, 1992).

Recently, the American Psychological Association (APA: 2001) has set forth additional guidelines for hypothesis testing in an effort to improve the current practices. They recommend, "take seriously the statistical power considerations associated with your test of hypotheses" (p. 24), and "to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your results section" (p. 25). While some researchers feel that statistical power considerations and effect size measures should be mandated (Cohen, 1988; Thompson, 1996), other researchers disagree (Frick, 1999; Levin & Robinson, 1999; Robinson & Levin, 1997).

But what is statistical power? What is an index of effect size? Power is the probability of avoiding a type II error or failing to reject a false null (Cohen, 1988). Power is affected by a number of factors including some factors that are directly controlled by the researcher and some that are not. Among the factors that are controlled by the researcher are sample size, statistical test, research design, and the alpha level. The factor that is not controlled by the researcher is the effect size. An effect size is a measure of magnitude of difference. While significance testing tells the researcher that there is a significant difference, effect size lets the researcher know how much of a difference. Although power is affected by sample size, effect size indices are not. This allows comparisons between different studies' effect sizes, which are based on different sample sizes (Clark-Carter, 1997).

Several papers have been written tracing the beginning and historical developments of effect size measures (Dwyer, 1974; Glass & Hakstian, 1969; Maxwell, Camp, & Arvey, 1981; Huberty, 2002; Richardson, 1996).

Researchers have also presented articles offering admonitory notes on the interpretation of effect size measures (Mitchell & Hartmann, 1981; Muray & Dosser, 1987; O'Grady, 1982; Sechrest & Yeaton, 1982; Strube, 1988). Despite the efforts of many to provide information about effect size indices researchers fail to utilize them (Keselman et al., 1998; McNamara, 1978). Recently, Olejnik and Algina (2000) wrote an article, which provides descriptions and formulas for comparative studies effect size measures, and a symposium discussing the current status of effect size measures made recommendations to the field (Elmore, 2001; Huberty 2001).

While the history of effect size measures have been traced, statistical power continues to be ignored in the research whether it be tracing its roots or applying it to current research designs. Researchers have asked if statistical power studies have had any effect on the statistical power of the current research. Some have replied no (Sedlmeier & Gigerenzer, 1989). Cohen first assessed the statistical power of psychological research in 1962. Rossi (1990) followed up Cohen's study by examining psychological research twenty years later.

With all the current awareness of the issues surrounding statistical power and effect size measures, it would be reasonable to assume that both have impacted the current research. The present study is an opportunity to assess the present status of psychological research and ask, "Are we improving or just staying the same."

## Research Hypotheses

The general objective of this study is to replicate the Cohen (1962) and Rossi (1990) power surveys and to complete an effect size survey of the same articles. The research hypotheses are as follows:

$H_1$. As in Rossi's (1990) replication of Cohen's study, a slight increase in statistical power is anticipated.

$H_2$. In respect to the effect size findings, a medium effect size is anticipated. The medium effect size described by Cohen (1992) with the intentions that it would be "visible to the naked eye of a careful observer."

$H_3$. Low statistical power and infrequent reporting of effect size measures will be prevalent.

## Study One:
## Assessing the Statistical Power
## Of Three Psychological Journals

Cohen's (1962) original power survey has set the guidelines for the procedures used in statistical power surveys. His original systematic approach involved surveying all of the articles published in Journal of Abnormal and Social Psychology for the year 1960. He only included articles in which statistical tests were conducted and tests which examined the primary hypotheses. Rossi's (1990) follow-up power survey mimicked Cohen's procedure and analyzed all of the articles published in the Journal of

Abnormal Psychology, 1982, vol. 91; the Journal of Consulting and Clinical Psychology, 1982, vol. 50; and the Journal of Personality and Social Psychology, 1982, vol. 42. Both power surveys had difficulty determining the power of all the potential articles. Some articles did not report one or more of the required factors to calculate power (sample size, alpha level, and effect size). (Additional information considering the required elements for calculating statistical power by statistical analysis choice is presented in table 4.1).

---

Insert Table 4.1 here

---

Method

Selection of Articles: All articles published in the Journal of Consulting and Clinical Psychology, 2000, vol. 68, the Journal of Personality and Social Psychology, 2000, vol. 78, and the Journal of Abnormal Psychology, 2000, vol. 109 were eligible for inclusion.

Articles were included in the analyses if:

1. Statistical tests were reported
2. Power can be calculated for the statistical tests that were reported.

Selection of Statistical tests: In agreement with previously run power surveys, a distinction was made between major and peripheral statistical tests. In this research study, major tests examine the research hypotheses of the study. Peripheral tests will examine research questions that are not the main hypotheses of the study.

Calculating Statistical Power: The PASS software (NCSS: 2001) will be used to determine power level in conjunction with power programs developed by Rossi (1990).

The following decision rules were employed:

1. Where a between-subjects design was utilized and subsamples sizes were not available, the subsamples were treated as equal. By assuming equal subsamples, maximum power was given for the research design.
2. Two-tailed tests with an alpha level of .05 were assumed for all tests.
3. Where a within-subjects design was employed in t-tests and ANOVA, the correlation was assumed to be .5 as recommended by Cohen (1988) and Lipsey (1990).
4. When ANCOVA was used, the correlation was assumed to be .5 as recommended by Cohen (1988) and Lipsey (1990).
5. Statistical tests in which many assumptions were to be made and little research evidence supporting these assumptions were excluded (i.e. survival analysis, path analysis, logistic regression, hierarchical linear modeling, etc.). Table 4.2 presents the

frequencies of all statistical tests (as described by the researchers themselves) utilized to address primary hypotheses.

---

Insert Table 4.2 here

---

6. For methods in which small, medium, and large effect sizes have not been operationally defined, power will be calculated but not included in general results (i.e., logistic regression: small, medium, and large effect size values for the odds ratio have not be published).

Results

Description of Articles

A pool of 287 potential articles were examined for inclusion in the study (75 in the Journal of Personality and Social Psychology; JPSP, 119 in the Journal of Consulting and Clinical Psychology; JCCP, and 93 in the Journal of Abnormal Psychology; JAP). A total of 157 articles were included in the statistical power survey examining small, medium, and large effect sizes. An additional 21 articles were examined for the power of the structural equation models and meta-analytic surveys presented. A total of 10 articles did not report statistics at all (3 in JPSP and 7 in JCCP), and 99 contained statistics for the primary hypothesis for which power was not determined (i.e. factor analysis, logistic regression, hierarchical linear modeling). Table 4.3 presents a description of the articles included in the assessment of small, medium, and large effect sizes.

---

Insert Table 4.3 here

---

Assessing Small, Medium, and Large

Statistical power was calculated for the remaining 157 articles (45 in JPSP, 57 in JCCP, and 55 in JAP). The total number of statistical tests for which power was calculated for was 2,747 with a total of 22,705 power calculations (power was calculated for effects beyond the normal small, medium, and large effect sizes). The frequency of the statistical tests included in the statistical power survey are presented in table 4.4. Due to the limitation of current statistical power survey techniques, the sample is dominated by "traditional" statistical tests such as the t-test, Pearson's correlation, and the ANOVA.

---

Insert Table 4.4 here

---

Because the number of statistical tests included in an article varied greatly (from zero to over 100), the article was used as the unit of analysis. This equalized all articles and allowed them all to contribute equally to the power survey results. The determination for small, medium, and large was assessed by averaging across statistical tests (for which small, medium, and large was available). Statistical power for small, medium, and large effects was calculated using both Cohen's (1962) original estimates and his later (1977) definitions.

Comparing Cohen '62, Rossi '90, and the Current Study

Statistical power estimates for the 1962 definitions are presented in table 4.5. A slight increase in statistical power can be noted across all three levels. Each increase is statistically significant, small: $F$ (2, 445) = 6.681, $p<$ .01, two-tailed, $\eta^2 = .029$, medium: $F$ (2, 445) = 16.067, $p<$ .01, two-tailed, $\eta^2 =$ .067, and large: $F$ (2, 445) = 5.731, $p<$ .01, two-tailed, $\eta^2 = .025$. Follow-up t-test results are reported in table 4.6 and show significant changes in medium effects across the three studies. Statistical power is not statistically different for small and large effects between the 1990 study and the current study. Table 4.7 compares the percent of studies with statistical power less than .50 and .80 in the 1990 study and the current study.

---

Insert Tables 4.5, 4.6, & 4.7 here

---

Statistical power estimates for the 1977 definitions are presented in table 4.8. A slight increase in statistical power can be noted across all three levels of effect size. Each increase is statistically significant, small: $t$ (376) = 3.303, $p<$ .01, two-tailed, $\eta^2 = .028$, medium: $t$ (376) = 3.768, $p<$ .01, two-tailed,

$\eta^2 = .036$, and large: $t\ (376) = 3.883$, $p < .01$, two-tailed, $\eta^2 = .038$. Table 4.9 compares the statistical power of the three journals using the 1962 and 1977 effect size definitions. Table 4.10 and figures 4.1 to figure 4.7 present the average power for a variety of effect size levels by statistical test. The figures and table show a rapid increase in observed power for some statistics and a slow increase for other statistics.

---

Insert Tables 4.8, 4.9, 4.10, & Figures 4.1 through 4.7 here

---

Results of SEM and meta-analysis power

Of the 71 structural equation models presented in the articles, 70 reported the required information for computing statistical power. Power was estimated for exact, close, and loose fitting models (RMSEAs of .05, .075, and .10 respectively). Power was determined to be .055 for exact fitting models, .518 for close fitting models, and .697 for loose fitting models. Of the 5 meta-analytic surveys, none of the surveys provided sufficient information for the calculation of power.

## Discussion

As the aforementioned tables exhibit, statistical power has continued to increase slightly from Cohen's (1962) original survey and Rossi's (1990) follow up survey. While the change is statistically significant, is it meaningful? The effect sizes are fairly small (eta-squared around .03). The statistical power of for small effects on the whole is poor (mean = .206) while the average power for large effects (mean = .884) is sufficient. While the power for medium effects is increasing, it is still not sufficient. Currently, it appears that research on average has about a 65% chance of detecting a medium effect. In Cohen's study, he noted that with respect to medium effect sizes, 57% of all studies had a power of .50. Rossi noted that 38% of studies in his survey had power less than .50. In the current study, the number continues to drop (32.5%). At the current rate, it will take close to 120 years for all studies to have sufficient statistical power to detect medium effects.

Cohen (1988) recommended a statistical power level of .80 (or 80% of the time correctly rejecting a false null). Current research designs are only sufficient in approaching this suggestion when examining large effect sizes. Unfortunately, once again these findings are not encouraging for psychological scientists. Researchers seem to still be failing in terms of statistical power.

Most power surveys are based on Cohen's notion that the medium effect size is most prevalent in psychological research. The next study examines the effect sizes of the previously mentioned articles.

## Study Two:
## Assessing the Effects Size Measures
## Of Three Psychological Journals

While statistical power surveys are easily accessible, surveys of effect sizes are difficult to uncover. In the previous chapters, effect size surveys have been noted. Each survey examined the average effect size for a journal or topic area. The current survey is designed to assess the average effect size indices for a variety of measures. To be included for examination are: eta-squared, Cohen's *d*, *r* (when reported as an effect size measure), *r*-squared, and odds ratios.

## Method

Selection of Articles: All articles published in the Journal of Consulting and Clinical Psychology, 2000, vol. 68, the Journal of Personality and Social Psychology, 2000, vol. 78, and the Journal of Abnormal Psychology, 2000, vol. 109 were eligible for inclusion.

Calculating Effect Size : The PASS software (NCSS: 2001) will be used to determine effect size in conjunction with power programs developed by Rossi (1988, 1990). Additional calculations will be done by hand using the following formulas (in addition to the formulas noted in previous chapters):

For the one-way ANOVA:

$$\eta^2 = (k\text{-}1)F/((k\text{-}1)F + \text{N} - k) \quad (4.1)$$

where k is the number of groups. For the t-test:

$$\eta^2 = t^2/(t^2 + \text{df}) \quad (4.2)$$

where df is the degrees of freedom. For factorial designs:

(4.3) $$\eta^2 = SS_{effect}/SS_{total}$$

Articles will be included in the analyses if:

1. Statistical tests were reported
2. Effect Size can be calculated for the statistical tests that were reported

Selection of Statistical tests: A distinction was made between major and peripheral statistical tests. In this research study, major tests examine the research hypotheses of the study. Peripheral tests will examine research questions that are not the main hypotheses of the study. Only effect sizes for major tests will be examined.

The following decision rules were employed:

1. When sufficient information was not reported (i.e. no means, no standard deviation, F test result, t test result), effect size measure was not calculated.
2. When effect size measures were reported, values for major tests were recorded.
3. For statistical tests in which widely accepted effect size measures are not available, effect size will not be calculated (i.e. SEM).

## Results

### Description of Articles

A pool of 287 potential articles were examined for inclusion in the study (75 in the Journal of Personality and Social Psychology; JPSP, 119 in the Journal of Consulting and Clinical Psychology; JCCP, and 93 in the

Journal of Abnormal Psychology; JAP). A total of 103 articles were included in the examination of effect sizes (23 in JPSP; 42 in JCCP; 38 in JAP). Articles were excluded if the statistical tests did not lend themselves to effect size calculation or sufficient information was not present. Because the number of statistical tests included in an article varied greatly (from zero to over 100), the article was used as the unit of analysis. This equalized all articles and allowed them all to contribute equally to the effect size survey results.

Effect Size Survey Results

A total of 42 articles either presented information concerning eta-squared or provided information in which enabled calculation of the measure. A total of 1,661 effect size measures were examined. A mean effect size of .194 and a median of .145 were determined for eta-squared. Of the studies that reported r-squared, a mean effect size of .259 and a median of .120 was discovered. Of the articles which utilized r as an effect size measure, .343 was the mean and .325 was the median. Researchers who calculated an odds ratio had 3.21 and 1.67 for their mean and median respectively. In the articles utilizing Cohen's d for their effect size measure, the mean was .672 and the median was .476. Table 4.11 presents a summary of the effect size findings.

---

Insert Table 4.11 here

---

## Discussion

The results of the effect size survey are encouraging. While no definition of a small, medium, and large effect exists for the odds ratio, Cohen (1988) did offer definitions for Cohen's *d*, Pearson's r, and eta-squared. It seems that on average a medium to large effect is prevalent in the research (when examining the mean and medium effect size).

These findings need to be interpreted with caution. Currently the APA (2001) only suggests reporting effect sizes for statistically significant findings. This suggestion in itself biases the reporting of effect sizes. Of the entire selection of articles examined that reported effect sizes, effect sizes were not reported for statistical tests which were not significant. On the whole, even when effect size measures were reported, they were not interpreted in terms of the research being examined.

## General Discussion

The statistical power survey exhibited that the statistical power of current psychological research is slowly increasing while the effect size survey confirmed Cohen's (1988) previous assertion that the most prevalent effect size in psychology is the medium effect size. The disappointing finding of the power survey is that the majority of psychological research still has insufficient statistical power to assess the magnitude of effects that they are examining.

In an effort to assess whether the fault of the use of statistical power lay with the current introduction to statistics textbooks, a quick examination of 11 textbooks published from 1999-2001 was run. Each textbook was examined by reading the hypothesis testing sections and checking the index for references on statistical power and effect size measures. Almost half of the textbooks (5 out of 11) address statistical power. Additional textbooks addressed statistical power by only giving the definition and the chart used in chapter 1. Surprisingly, only a few textbooks discussed effect size measures with any substance (refer to table 4.12 for a summary).

---

Insert Table 4.12 here

---

It would seem that the field is at least moving towards the utilization of statistical power while effect size indexes are not getting as much attention. According the current study though, 30.7% of the studies examined reported using effect size measures while only 7% reported examining statistical power.

Suggestions for improvement

The analysis of advanced statistical methods such as logistic regression, HLM, SEM, and path analysis need to be addressed (especially when examining power surveys across topics). Statistical power survey methods need to be developed to examine longitudinal methods and other common statistical analysis techniques. The procedures of ANCOVA and repeated measures cannot be addressed across topic areas without causing detriment to some areas and benefiting others. It is difficult to make assumptions about these techniques that fit psychological research as a whole. Another suggestion for future statistical power surveys and effect size surveys would be to break down the articles by topic (e.g., PTSD, eating disorders, addiction).

After examining the data in the power survey and effect size survey, it seems evident that it is easiest for those who are running the analyses to compute statistical power and effect size indices. It is very hard for post publication calculation. To further psychological as a science, it is paramount that these calculations precede publication and are reported.

References

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the Chi-square test. Journal of the American Statistical Association, 33, 526-536.

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, revisited. Journal of Experimental Education, 61, 287-292.

Chow, S. L. (1998). Precis of statistical significance: Rationale, validity, and utility. Behavioral and Brain Sciences, 21, 169-239.

Clark-Carter, D. (1997). The account taken of statistical power in research published in the British Journal of Psychology. British Journal of Psychology, 88, 71-83.

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

Cohen, J. (1988). Statistical power for the behavioral sciences. (2nd Ed.) New York: Academic Press.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

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

Dwyer, J. H. (1974). Analysis of variance and the magnitude of effects. Psychological Bulletin, 81, 731-737.

Elmore, P. B., & Rotou, O. (2001). A primer on basic effect size concepts. Presented at the Annual meeting of the American Educational Research Association, Seattle, WA.

Frick, R. W. (1999). Defending the status quo. Theory & Psychology, 9, 183-189.

Glass, G. V., & Hakstian, A. R. (1969). Measures of association in comparative experiments: Their development and interpretation. American Educational Research Journal, 5, 3-8.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). What if there were no significance tests? Mahwah, NJ: Erlbaum.

Hays, W. L. (1963). Statistics for psychologists. New York: Holt, Rinehart, & Winston.

Hintze, J. L. (2000). PASS 2000 user's guide. Kaysville, Utah: Number Cruncher Statistical Systems.

Hogben, L. (1957). Statistical theory: The relationship of probability, credibility, and error. An examination of the contemporary crisis in statistical theory from a behaviorist viewpoint. London: Allen & Unwin.

Huberty, C. J. (2001, April) A History of Effect Size Indices. Presented at the Annual meeting of the American Educational Research Association, Seattle, WA.

Huberty, C. J. (2002) A history of effect size indices. Educational and Psychological Measurement, 62, 227-240.

Hunter, J. E. (1997). Needed: A ban on the significance test. Psychological Science, 8, 3-7.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olenjnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of educational researchers: An anlysis of there ANOVA, MANOVA, and ANCOVA analysis. Review of Educational Research, 68, 350-386.

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

Levin, J. R., & Robinson, D. H. (1999). Further reflections on hypothesis testing and editorial policy for primary research journals. Educational Psychology Review, 11, 143-155.

Lipsey, M. W. (1990) Design Sensitivity Statistical power for Experimental Research. London: Sage Publications, Inc.

Maxwell, S. E., Camp, J. C., & Arvey, R. D. (1981). Measures of strength of association: A comparative examination. Journal of Applied Psychology, 66, 525-534.

McNamara, J. F. (1978). Practical significance and statistical models. Educational Administration Quarterly, 14, 48-63.

Mitchell, C., & Hartmann, D. P. (1981). A cautionary note on the use of omega-squared to evaluate the effectiveness of behavioral treatments. Behavioral Assessment, 3, 93-100.

Morrison, D.E., & Henkle, R. E. (Eds.). (1970). The significance test controversy. Chicago: Aldine.

Murray, L. W., & Dosser, D. A., Jr. (1987). How significant is a significant difference? Problems with the measurement of magnitude of effect. Journal of Counseling Psychology, 34, 68-72.

O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. Psychological Bulletin, 92, 766-777.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. Contemporary Educational Psychology, 25, 241-286.

Richardson, J. T. E. (1996). Measures of effect size. Behavioral Research Methods, Instruments, & Computers, 28, 12-22.

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

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? Journal of Consulting and Clinical Psychology, 58, 646-656.

Schmidt, F. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. American Psychologist, 47, 1173-1181.

Sechrest, L., & Yeaton, W. H. (1982). Magnitudes of experimental effects in social science research. Evaluation Review, 6, 579-600.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? Psychological Bulletin, 105, 309-316.

Strube, M. J. (1988). Some comments on the use of magnitudes-of-effects estimates. Journal of Counseling Research, 35, 342-345.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researchers, 25, 26-30.

Thompson, B. (1997a). Editorial policies regarding statistical significance tests: Further comments. Educational Researcher, 26, 29-32.

Thompson, B. (1997b). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. Educational & Psychological Measurement, 57, 5-19.

Thompson, B. (1999). Why "encouraging" effect size reporting is not working: The etiology of researcher resistance to changing practices. The Journal of Psychology, 133, 133-140.

Thompson, B. (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? Journal of Counseling and Development, 80, 64-71.

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

Table 4.1

Required Information for Determining Power by Statistical Test

| Statistic | Required Elements | | | | | | |
|---|---|---|---|---|---|---|---|
| | *df* | $H_a$ | #IV | #DV | Corr | #Grp | Other |
| *Univariate* | | | | | | | |
| Nonparametric | | | | | | | |
| Chi-square | x | x | | | | | |
| Fisher's Exact Test | x | x | | | | | Prob. |
| t test | | | | | | | |
| Independent | | x | | | | | |
| Dependent | | x | | | | | |
| Correlation | | | | | | | |
| Bivariate/Part/Partial | | x | | | | | |
| Classical Regression | | | | | | | |
| Hierarchical | | x | x | x | | | |
| Stepwise | | x | x | x | | | |
| Specialized Regression | | | | | | | |
| HLM | | x | x | x | | | ICC |
| Logistic | | x | x | x | | | OR |
| ANOVA (ANCOVA) | | | | | | | |
| Oneway | x | x | | | | | |
| Factorial | x | x | | | | | |
| Repeated Measures | x | x | | | | | auto |
| *Multivariate* | | | | | | | |
| Factor Analysis | | | *Analysis not appropriate* | | | | |
| EFA/ CFA | | | *Analysis not appropriate* | | | | |
| Cluster Analysis | | | *Analysis not appropriate* | | | | |
| Multidimensional Scaling | | | *Analysis not appropriate* | | | | |
| MANOVA (MACOVA) | | | | | | | |
| Oneway | x | x | | x | | x | |
| Factorial | x | x | | x | | x | |
| Post hoc Univariate | x | x | | x | | x | |
| Repeated Measures | x | x | | x | x | x | |
| Discriminant Analysis | x | x | | x | | x | |
| Canonical R Analysis | x | x | x | x | | | |
| SEM | x | | | | | | |

Note: Alpha level, sample size, and effect size are a requirement for all of the aforementioned statistics. $H_a$ is the directional nature of the hypotheses (one- versus two-tailed).

Table 4.2
Frequencies of Statistics in Potential Articles for Power and Effect Size

Survey

| Statistic | Frequency | JPSP | JCCP | JAP |
|---|---|---|---|---|
| **ANOVA** | | | | |
| One-way | 58 | 17 | 19 | 22 |
| 2-way | 5 | 2 | 3 | 0 |
| 2-way, repeated measures | 4 | 2 | 2 | 0 |
| 3-Way | 4 | 0 | 2 | 2 |
| Between-within | 1 | 1 | 0 | 0 |
| Factorial | 8 | 2 | 2 | 4 |
| Hierarchical | 1 | 0 | 0 | 1 |
| Meta-analytic | 2 | 1 | 1 | 0 |
| Mixed factorial | 3 | 1 | 1 | 1 |
| Mixed | 7 | 1 | 1 | 5 |
| Repeated measures | 33 | 9 | 9 | 15 |
| Simple Effects | 1 | 0 | 0 | 1 |
| Split Plot | 3 | 1 | 2 | 0 |
| **ANCOVA** | | | | |
| One-way | 25 | 9 | 12 | 4 |
| 3-way | 1 | 1 | 0 | 0 |
| repeated measures | 10 | 4 | 3 | 3 |
| **Bootstraping** | 2 | 1 | 1 | 0 |
| **Box Test** | 1 | 0 | 1 | 0 |
| **Chi-square** | 39 | 13 | 13 | 13 |
| **Cluster Analysis** | 2 | 1 | 1 | 0 |
| Hierarchical | 3 | 1 | 1 | 1 |
| **Confirmatory Factor Analysis** | 5 | 2 | 2 | 1 |
| **Correlation** | 54 | 17 | 17 | 20 |
| polychoric | 2 | 0 | 1 | 1 |
| **Cost Analysis** | 2 | 1 | 1 | 0 |
| **Cox Proportional Hazard Model** | 5 | 2 | 2 | 1 |
| **Cross Lagged Panel** | 2 | 0 | 0 | 2 |
| **Descriptives** | 11 | 7 | 4 | 0 |
| **DFA** | 5 | 1 | 1 | 3 |
| **Factor Analysis** | 8 | 3 | 3 | 2 |
| **Fisher's Exact Probability** | 7 | 3 | 3 | 1 |
| **Fisher's Exact Z Transformation** | 2 | 1 | 1 | 0 |
| **GEE** | 1 | 0 | 0 | 1 |
| **Growth Curve Analysis** | 2 | 1 | 1 | 0 |
| **Hierarchical Linear Modeling** | 7 | 2 | 2 | 3 |

| | | | | |
|---|---|---|---|---|
| **Hierarchical Multiple Linear Regression** | 1 | 0 | 1 | 0 |
| **Hierarchical Regression** | 4 | 0 | 1 | 3 |
| **HITMAX** | 1 | 1 | 0 | 0 |
| **Hosmer-Lemeshow** | 1 | 0 | 0 | 1 |
| **Item Response Theory** | 1 | 1 | 0 | 0 |
| **K-group split plot multivariate analysis** | 2 | 0 | 2 | 0 |
| **Kruskal-Wallis** | 2 | 0 | 1 | 1 |
| **Linear Regression** | | | | |
| Standard | 1 | 0 | 0 | 1 |
| Hierarchical | 3 | 0 | 0 | 3 |
| **Log Linear** | 1 | 1 | 0 | 0 |
| **Logistic Regression** | | | | |
| Standard | 20 | 0 | 12 | 8 |
| Hierarchical | 3 | 0 | 3 | 0 |
| Ordinal | 1 | 0 | 0 | 1 |
| Polytomous | 1 | 0 | 1 | 0 |
| Stepwise | 1 | 0 | 1 | 0 |
| **Longitudinal Modeling** (mixed) | 2 | 0 | 1 | 1 |
| **Longitudinal Random Regression** | 1 | 0 | 1 | 0 |
| **MAMBAC** | 1 | 0 | 0 | 1 |
| **Mann Whitney U** | 1 | 0 | 0 | 1 |
| **MANOVA** | | | | |
| One-way | 12 | 1 | 7 | 4 |
| Mixed model | 3 | 1 | 1 | 1 |
| Repeated measures | 5 | 1 | 3 | 1 |
| Within subjects | 1 | 0 | 1 | 0 |
| **MANCOVA** | 2 | 0 | 2 | 0 |
| **MAXCOV** | 2 | 0 | 0 | 2 |
| **Mediational Analysis** | 8 | 3 | 4 | 1 |
| **Meta-analysis** | 5 | 2 | 3 | 0 |
| **Mixture Analysis** | 1 | 0 | 1 | 0 |
| **Multi Level Regression** | 1 | 0 | 0 | 1 |
| **Multidimensional Scaling** | 1 | 0 | 1 | 0 |
| **Multiple Regression** | | | | |
| Standard | 14 | 5 | 5 | 4 |
| Hierarchical | 10 | 1 | 6 | 3 |
| Stepwise | 3 | 0 | 2 | 1 |
| with Covariates | 1 | 0 | 1 | 0 |
| **Networks** | 1 | 0 | 1 | 0 |
| **Odds Ratio** | 2 | 0 | 2 | 0 |

| | | | | |
|---|---|---|---|---|
| **Path Analysis** | 8 | 4 | 3 | 1 |
| **Poisson Sampling Model** | 1 | 0 | 1 | 0 |
| **Principal Component Analysis** | 2 | 1 | 1 | 0 |
| **Profile Analysis** | 1 | 0 | 1 | 0 |
| **Proportional Hazards Analysis** | 1 | 0 | 0 | 1 |
| **Proportions** | 1 | 0 | 0 | 1 |
| **Random Coefficient Analysis** | 1 | 0 | 1 | 0 |
| **Receiver Operating Characteristic** | 1 | 0 | 1 | 0 |
| **Regression** | | | | |
| Regression | 9 | 1 | 7 | 1 |
| Moderated | 1 | 1 | 0 | 0 |
| Random Effects | 2 | 0 | 2 | 0 |
| Stepwise | 1 | 0 | 0 | 1 |
| **Reliable Change Index** | 1 | 0 | 1 | 0 |
| **ROC Analysis** | 1 | 0 | 1 | 0 |
| **Rom's Procedure** | 1 | 1 | 0 | 0 |
| **SEM** | 15 | 5 | 6 | 4 |
| **Survival Analysis** | 8 | 0 | 6 | 2 |
| **Taxometric** | 3 | 1 | 0 | 2 |
| **Trend Analysis** | 1 | 0 | 0 | 1 |
| **Trimmed Means** | 1 | 1 | 0 | 0 |
| **t-test** | 45 | 19 | 14 | 12 |
| **Yule's Y** | 1 | 0 | 1 | 0 |
| **z-score** | 1 | 0 | 0 | 1 |

Table 4.3

Description of Articles Included in Current Study

| | JPSP | JCCP | JAP | Total |
|---|---|---|---|---|
| **First Author** | | | | |
| Female | 13 (28.9) | 23 (40.4) | 23 (40.0) | 58 |
| Male | 26 (57.8) | 33 (57.9) | 25 (45.5) | 84 |
| Unknown | 6 (13.3) | 1 (1.8) | 8 (15.1) | 15 |
| **Mean Number of Authors** | 2.89 | 4.33 | 4.00 | 3.80 |
| **Median Number of Authors** | 3.00 | 4.00 | 4.00 | 3.00 |
| **Median Sample Size** (across studies) | 296 | 92 | 81 | 146 |
| Females | 133 | 36 | 39 | 57 |
| Males | 102 | 27.5 | 30 | 40 |
| **Mean Number of Experiments** | 3.0 | 1.02 | 1.13 | 1.62 |
| **Median Number of Experiments** | 3 | 1 | 1 | 1 |
| **Grant Funded** | | | | |
| Yes | 32 (71.1) | 46 (80.7) | 42 (76.4) | 120 |
| No | 13 (28.9) | 11 (19.3) | 11 (20.0) | 35 |
| **Presented at a Conference** | | | | |
| Yes | 8 (17.8) | 5 (8.8) | 6 (10.9) | 19 |
| No | 37 (82.2) | 52 (91.2) | 47 (85.5) | 136 |
| **Student Thesis or Dissertation** | | | | |
| Yes | 4 (8.9) | 5 (8.8) | 8 (14.5) | 17 |
| No | 41 (91.1) | 52 (91.2) | 45 (81.8) | 138 |
| **Used Alpha Correction Methods** | | | | |
| Yes | 1 (2.2) | 9 (15.8) | 7 (12.7) | 17 |
| No | 44 (97.8) | 47 (82.5) | 48 (87.3) | 139 |
| **Number of Articles without CI** | 44 (97.8) | 53 (93.0) | 50 (90.9) | 147 |
| **Number of Articles mentioning Power** | 1 (2.2) | 1 (1.8) | 4 (7.7) | 6 |
| **Calculated Effect Size** | | | | |
| Yes | 11 (24.4) | 23 (40.4) | 16 (29.1) | 50 |
| No | 34 (75.6) | 33 (57.9) | 39 (70.9) | 106 |
| **Number of Brief Reports** | 0 | 10 (17.5) | 12 (21.8) | 22 |

Note: Percentages in the parentheses.

Table 4.4

Frequency Distribution of Statistical Tests included in Power Survey

| Statistical Test | Frequency | Proportion |
| --- | --- | --- |
| Pearson Correlation | 1429 | .520 |
| ANOVA | 753 | .274 |
| t-test | 250 | .091 |
| ANCOVA | 120 | .043 |
| Chi-square | 87 | .031 |
| Multiple Regression | 56 | .020 |
| MANOVA | 52 | .018 |

Table 4.5

Average Power of Current Study in Comparison to Rossi (1990) and Cohen (1962) Based on the 1962 Effect Size Definitions

| Effect Size | Mean | SD | Median |
|---|---|---|---|
| 2002 Study (n=157) | | | |
| Small '62 | .278 | .238 | .186 |
| Medium '62 | .678 | .263 | .725 |
| Large '62 | .897 | .144 | .970 |
| Rossi (1990) (n=221) | | | |
| Small '62 | .240 | .171 | .184 |
| Medium '62 | .590 | .251 | .582 |
| Large '62 | .893 | .131 | .958 |
| Cohen (1962) (n=70) | | | |
| Small '62 | .18 | .08 | .17 |
| Medium '62 | .48 | .20 | .46 |
| Large '62 | .83 | .16 | .89 |

Table 4.6

Follow-up t-test Results Comparing the Current Study, Rossi (1990), and Cohen (1962)

| | Studies being compared | | Mean Difference | *t*-value | *p*<.05? |
|---|---|---|---|---|---|
| **Small Effect** | 2002 | 1990 | .038 | 1.807 | No |
| | | 1962 | .098 | 3.358 | Yes |
| | 1990 | 2002 | -.038 | 1.807 | No |
| | | 1962 | .060 | 2.836 | Yes |
| **Medium Effect** | 2002 | 1990 | .088 | 3.293 | Yes |
| | | 1962 | .198 | 5.614 | Yes |
| | 1990 | 2002 | -.088 | 3.293 | Yes |
| | | 1962 | .110 | 3.344 | Yes |
| **Large Effect** | 2002 | 1990 | .004 | .280 | No |
| | | 1962 | .067 | 3.217 | Yes |
| | 1990 | 2002 | -.004 | .280 | No |
| | | 1962 | .063 | 3.317 | Yes |

Table 4.7

Percentage of Studies with Power <.50 and <.80

| | Current Study | | 1990 Study | |
|---|---|---|---|---|
| Effect Size | <.50 | <.80 | <.50 | <.80 |
| Small | 91.1 | 96.2 | 98 | 99 |
| Medium | 32.5 | 68.2 | 46 | 84 |
| Large | 3.2 | 22.3 | 8 | 36 |

Table 4.8

Average Power of Current Study Compared to Rossi (1990) using 1977 Definitions

| Effect Size | Mean | SD | Median |
|---|---|---|---|
| 2002 Study (n=157) | | | |
| Small '77 | .206 | .194 | .139 |
| Medium '77 | .646 | .253 | .663 |
| Large '77 | .884 | .152 | .962 |
| Rossi (1990) (n=221) | | | |
| Small '77 | .153 | .117 | .121 |
| Medium '77 | .549 | .242 | .516 |
| Large '77 | .814 | .186 | .883 |

Table 4.9

Power and Effect Size Estimations by Journal

| | | Journal | | | |
|---|---|---|---|---|---|
| | | JPSP | JCCP | JAP | Total |
| **Power** | | | | | |
| Small '77 | Mean | .172 | .232 | .207 | .206 |
| | Median | .150 | .140 | .120 | .139 |
| | SD | .121 | .207 | .228 | .195 |
| Medium '77 | Mean | .626 | .686 | .620 | .646 |
| | Median | .654 | .709 | .600 | .663 |
| | SD | .221 | .263 | .266 | .253 |
| Large '77 | Mean | .885 | .897 | .869 | .884 |
| | Median | .950 | .970 | .955 | .962 |
| | SD | .129 | .160 | .162 | .152 |
| Small '62 | Mean | .261 | .300 | .270 | .278 |
| | Median | .176 | .235 | .180 | .186 |
| | SD | .207 | .257 | .244 | .238 |
| Medium '62 | Mean | .659 | .714 | .658 | .678 |
| | Median | .710 | .810 | .708 | .725 |
| | SD | .241 | .269 | .274 | .263 |
| Large '62 | Mean | .902 | .903 | .886 | .897 |
| | Median | .966 | .987 | .959 | .970 |
| | SD | .115 | .159 | .151 | .144 |
| **Effect Size** | | | | | |
| Eta-Squared | Mean | .195 | .220 | .181 | .194 |
| | Median | .192 | .179 | .142 | .154 |
| | SD | .121 | .167 | .154 | .147 |
| Odds Ratio | Mean | | 2.831 | 6.268 | 3.213 |
| | Median | | 2.209 | 6.268 | 2.209 |
| | SD | | 1.767 | 8.553 | 2.879 |
| R-squared | Mean | .274 | .242 | .261 | .259 |
| | Median | .273 | .150 | .247 | .176 |
| | SD | .229 | .193 | .210 | .204 |

Table 4.10

Average Power by Statistical Test

| | | Effect Size | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| t-test | d= | **0.1** | **0.2** | **0.25** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** | **1.0** |
| | | .08 | .18 | .24 | .31 | .44 | .56 | .66 | .75 | .81 | .87 | .91 |
| ANOVA | f= | **0.1** | **0.125** | **0.2** | **0.25** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** |
| | | .14 | .20 | .40 | .54 | .68 | .85 | .94 | .97 | .98 | .99 | .99 |
| ANCOVA | f= | **0.1** | **0.125** | **0.2** | **0.25** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | |
| | | .18 | .26 | .55 | .73 | .84 | .95 | .98 | .99 | .99 | .99 | |
| Correlation | r= | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | | | | | |
| | | .27 | .61 | .83 | .94 | .98 | .99 | | | | | |
| Multiple Regression | $R^2$= | **0** | **0.02** | **0.1** | **0.15** | **0.2** | **0.25** | **0.3** | **0.35** | **0.4** | | |
| | | .15 | .36 | .72 | .81 | .87 | .91 | .94 | .95 | .97 | | |
| Chi-square | w= | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** | **1.0** | |
| | | .47 | .66 | .81 | .91 | .97 | .99 | .99 | .99 | .99 | .99 | |
| MANOVA | $R^2$= | **0.02** | **0.05** | **0.1** | **0.15** | **0.2** | **0.25** | **0.3** | **0.35** | **0.4** | | |
| | | .54 | .81 | .91 | .93 | .93 | .94 | .94 | .94 | .94 | | |

Table 4.11

Description of the Average Effect Sizes in the 2002 Study

| | Effect Size Measure | | | | |
|---|---|---|---|---|---|
| | $\eta^2$ | $R^2$ | r | Odds Ratio | Cohen's d |
| Mean | .194 | .259 | .343 | 3.21 | .672 |
| Median | .145 | .120 | .325 | 1.67 | .476 |
| SD | .205 | .201 | .248 | 6.289 | .549 |
| Minimum | 0 | 0 | -.73 | 0 | -1.49 |
| Maximum | 1.02 | .944 | .89 | 55.38 | 3.88 |
| Percentiles | | | | | |
| 25th | .00933 | .00798 | .165 | 1.423 | .306 |
| 50th | .145 | .120 | .325 | 1.67 | .476 |
| 75th | .290 | .443 | .460 | 4.015 | .907 |
| 80th | .312 | .524 | .606 | 4.998 | .929 |
| Number of Studies | 42 | 30 | 7 | 18 | 18 |
| Number of Tests | 416 | 322 | 112 | 404 | 407 |

Table 4.12

Survey of Introduction to Statistic Textbooks published in the 2000 and the Coverage of Power and Effect Size Measures

| Textbook | Publisher | Power | Effect Size |
|---|---|---|---|
| Applied Statistics for the Behavioral Sciences | Houghton Mifflin | Yes | Yes |
| Essentials of Statistics | Brooks/Cole | No | No |
| Everyday Statistical Reasoning | Wadsworth | No | No |
| Statistical Analysis | Radius Press | Briefly | No |
| Statistics and Data Analysis | McGraw Hill | Yes | Yes |
| Statistics For People who (think they) Hate Statistics | Sage | Briefly | No |
| Statistics for the Behavioral and Social Sciences | Prentice Hall | Yes | Yes |
| Statistics for the Behavioral Sciences | Wadsworth | Yes | No |
| Statistics with Confidence | Sage | Yes | Yes |
| Student Friendly Statistics | Prentice Hall | Briefly | No |
| The Cartoon Guide to Statistics | Harper Perennial | Briefly | No |

Figure 4.1

Power curve for t test statistics included in the present study

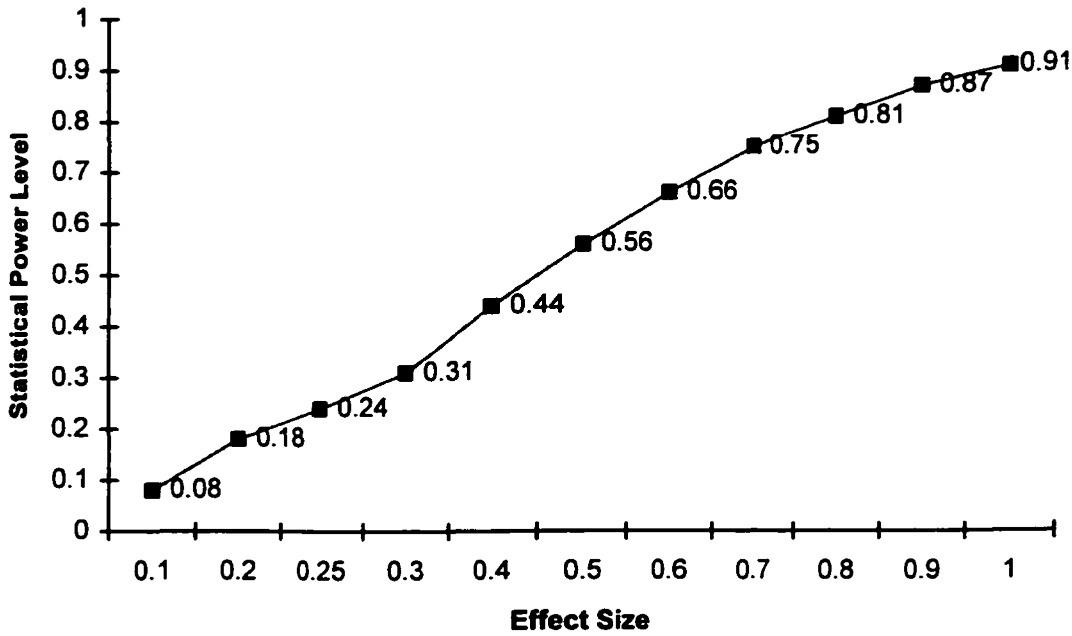


Figure 4.2

Power curve for ANOVA statistics included in the present study

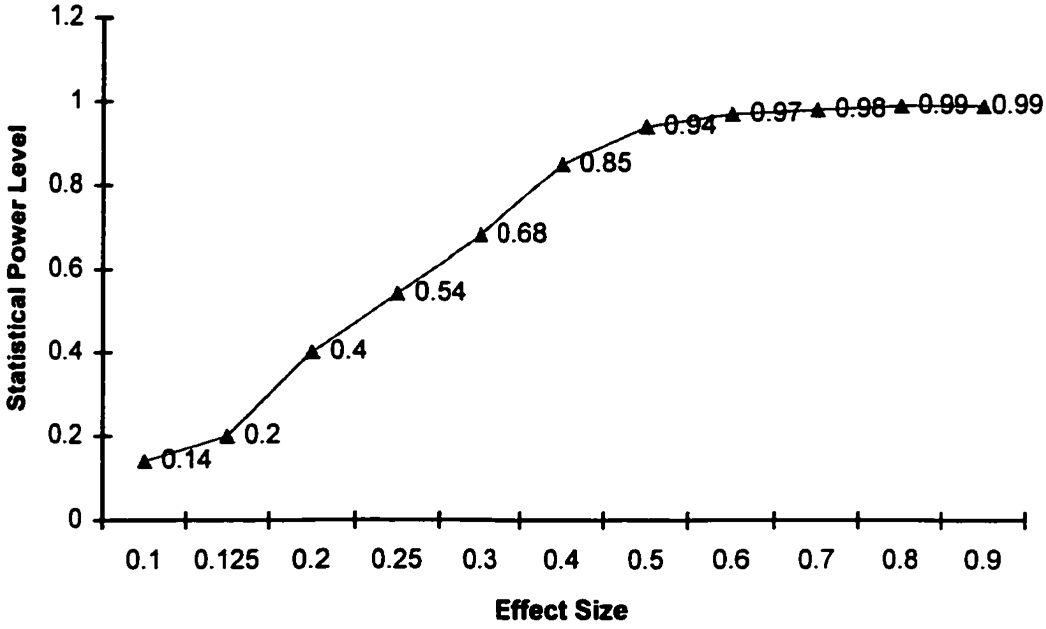


Figure 4.3

Power curve for ANCOVA statistics included in the present study

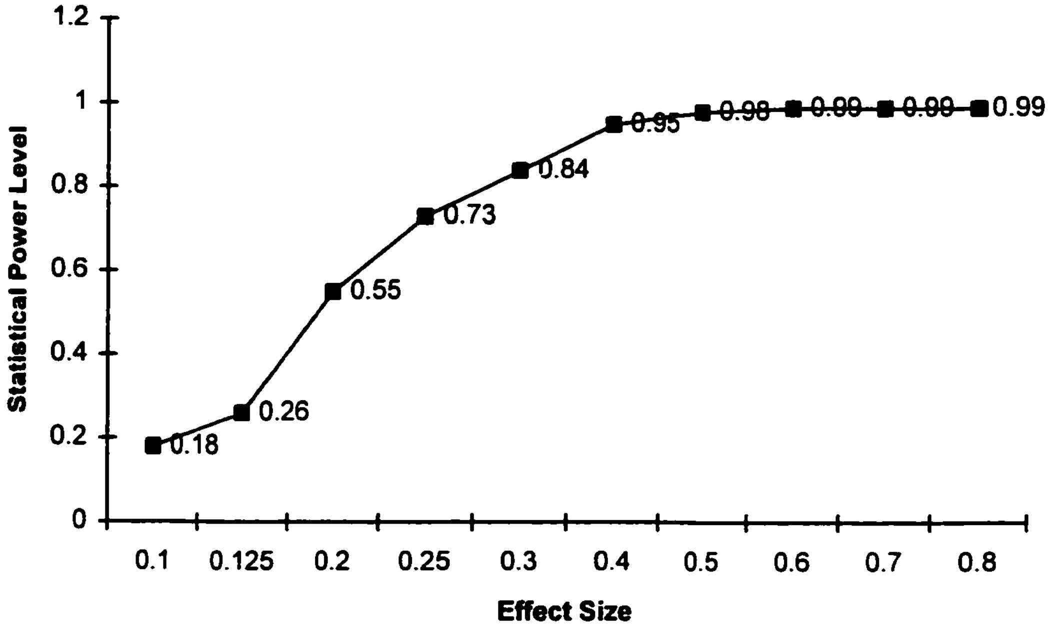


Figure 4.4

Power curve for correlation statistics included in the present study

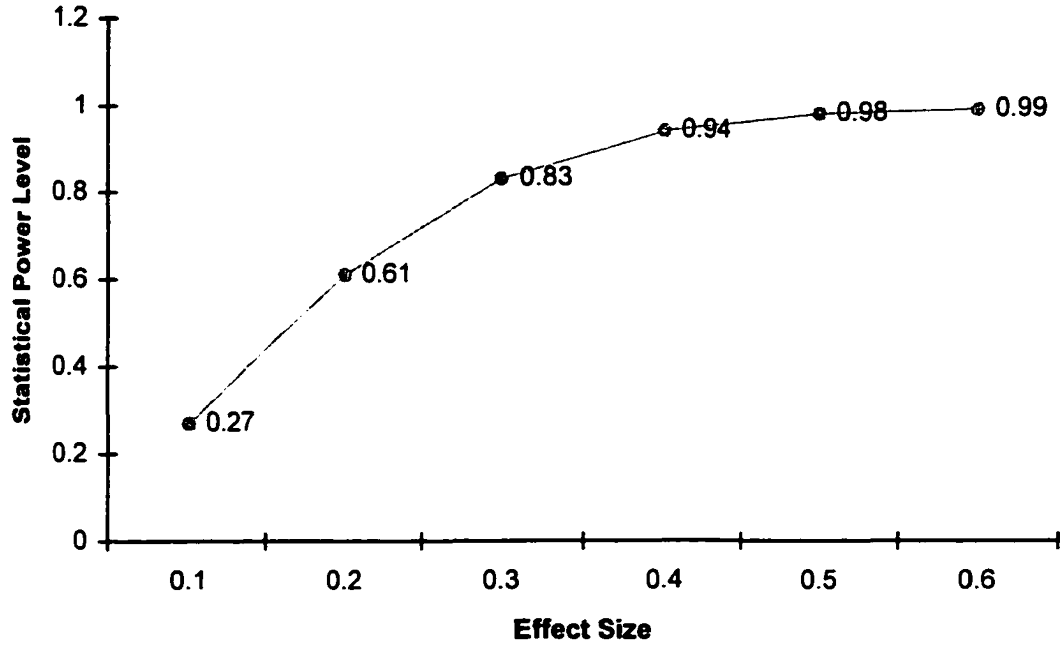


Figure 4.5

Power curve for multiple regression statistics included in the present study

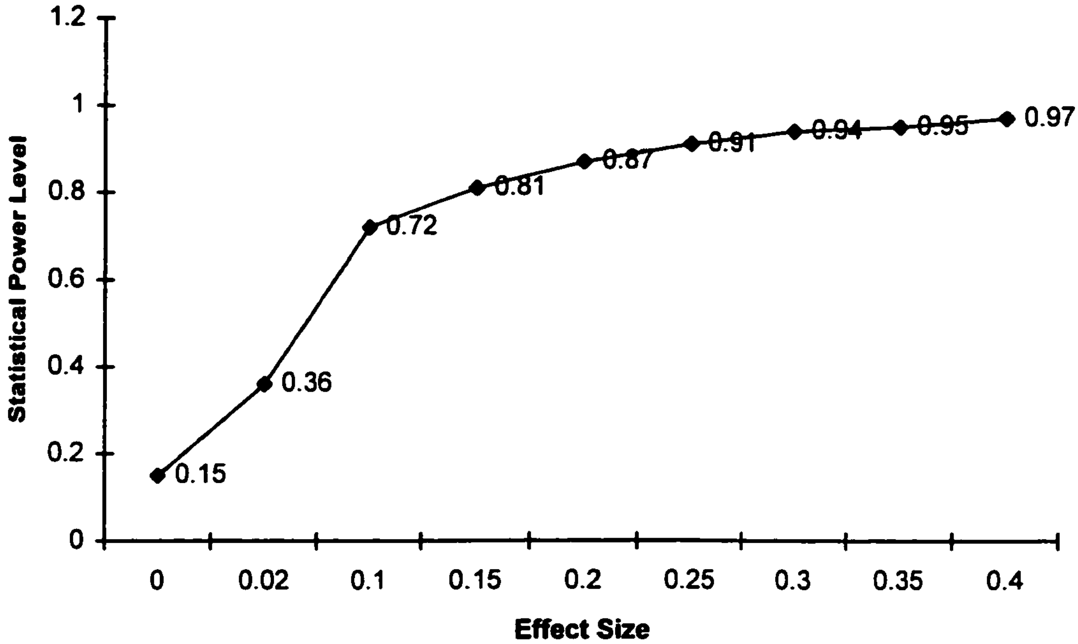


Figure 4.6

Power curve for chi-square statistics included in the present study

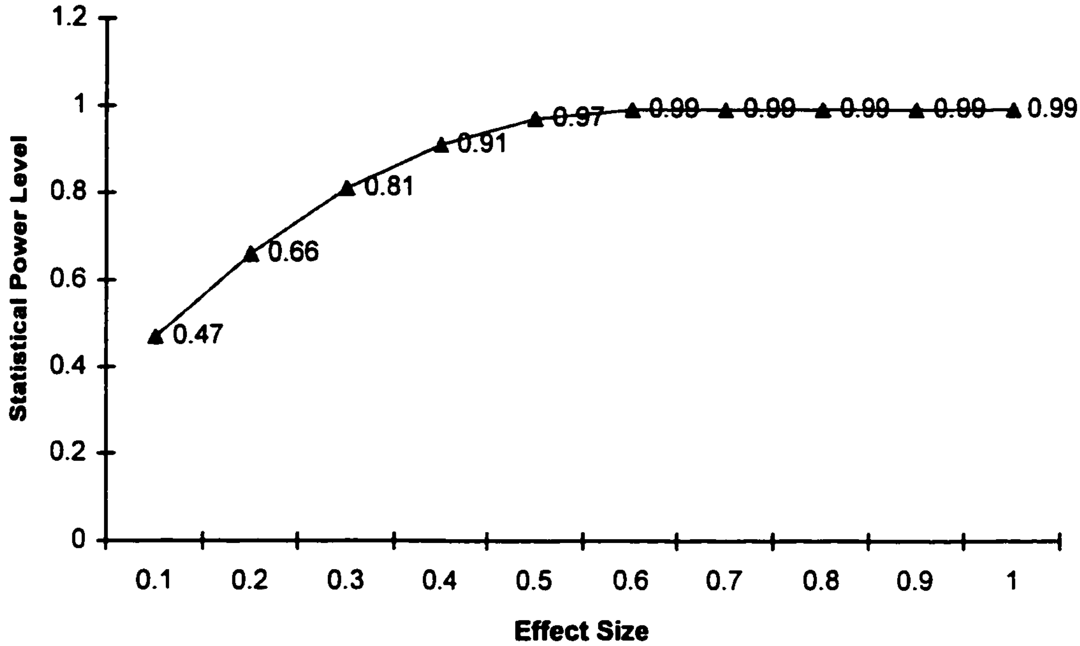
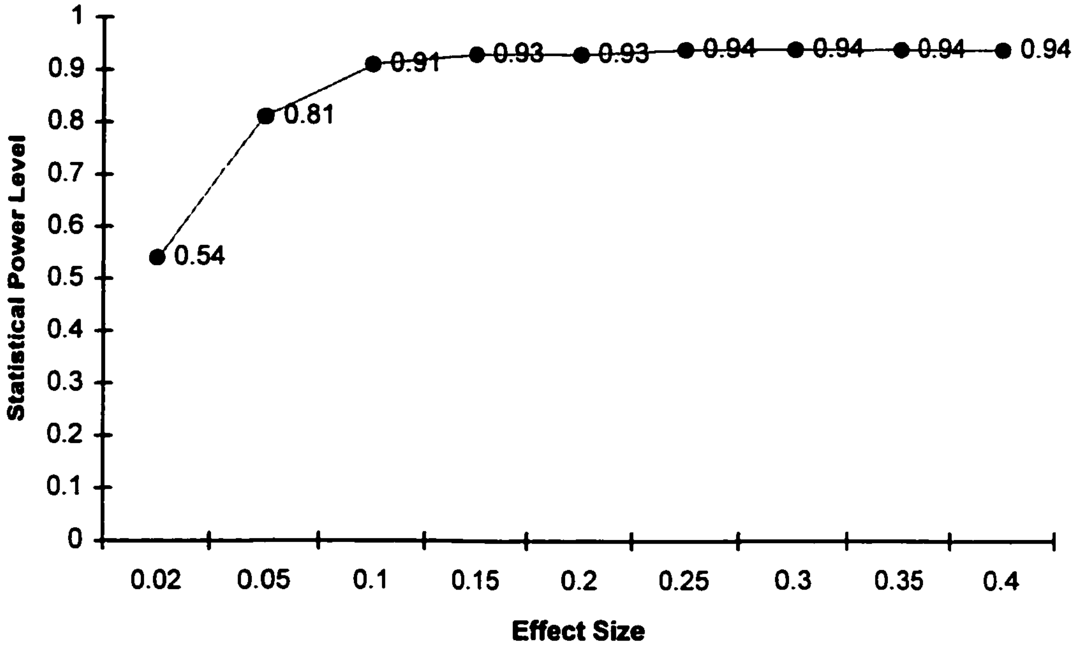


Figure 4.7

Power curve for MANOVA statistics included in the present study



*Where is the wisdom we have lost in knowledge?*
*Where is the knowledge we have lost in information?*
*T. S. Eliot (1888 - 1965)*

# Chapter 5:

## Summary of Findings

Statistical power analysis and effect size indices have been formulated in the past, applied in the current literature (on a limited basis), and will continue to evolve in future literature and research practices. While the practices of the past are being used in the current literature, for the both techniques to evolve to the next level, researchers need to conceptualize and utilize these techniques differently.

### Foundations of Statistical Power and Effect Size

Statistical power and effect size indices have a fairly short history. Both have evolved from their earlier forms to the current status of complementing null hypothesis significance testing practices. Statistical power is an extension of Neyman and Pearson's (1928a, 1928b, 1933a, 1933b) initial work in the area of significance testing. Some initial effect size indices predate Neyman and Pearson's work, for example Yule's (1900) work examining how different populations are or the magnitude of differences.

The pioneer in terms of statistical power in current psychological research is Jacob Cohen (1962, 1988). His initial statistical power survey set the guidelines for future surveys. Cohen (1962) established that power would be calculated with respect to three effect sizes (small, medium, and large), with respect to an alpha level of .05, assuming a two-tailed test, for statistical tests that are central to the primary hypotheses, and for statistical tests in which statistical power techniques are available. Cohen also established that the article would be the unit of analysis because some articles had multiple statistical tests while others had only a few tests.

Many researchers followed in Cohen's steps and surveyed psychological, communication, marketing, and many other types of journals to assess the statistical power in those fields. Other researchers extended Cohen's research and examined the statistical power of a specific research interest (i.e. research on the Rorschach, and research on psychotherapy treatment homework). These statistical power surveys were focused on the research interest and able to cross a spectrum of research journals.

Effect size research has not followed in the same path. Fewer researchers are concerned with surveying the literature to determine the status of effect sizes. Most researchers accept Cohen's initial recommendations for "small," "medium," and "large" effect sizes with out question while Cohen admitted that his definitions were based on his own experiences—no research. The few effect size surveys available do not address Cohen's definition nor posit their own definitions for small, medium, and large. It might seem logical that small, medium, and large may differ by research area or research interest, yet the research to confirm this assertion is not available.

Changing the whole system

One suggestion to problems with current research practices is to remove significance testing entirely. Statistical power and effect size measures were developed to complement the null hypothesis significance testing (NHST) message. If NHST was to be disbanded, it is proper to assume that at least statistical power would cease to exist. In terms of NHST, statistical power is the ability to discover a significant effect when a significant effect is present. If researchers are no longer testing for significant effect, determining statistical power is, in essence, meaningless.

While statistical power would vanish, effect size indices could still complement research findings. Instead of reporting that groups or treatments are different, effect size measures would indicate how much of a difference is present. Effect size measures would not state if the differences were "statistically" different, but indicate how much the groups differed.

Even as the disbandment of NHST has been suggested, it has not be broadly accepted, nor enforced. Despite the controversy, statistical power and effect size practices continue to evolve. Currently, methods are available to calculate statistical power for a variety of statistical techniques (i.e. correlation, t-test, ANOVA, ANCOVA, MANOVA, and SEM) and to calculate a variety of effect size measures (i.e. eta-squared, omega-squared, intraclass correlation, odds ratio, epsilon-squared, etc.). A plethora of statistical packages are available to calculate both statistical power (i.e.

GPOWER, SAS, and PASS) and effect size indices (i.e. SAS, SPSS, and EXCEL).

However, researchers fail to utilize and interpret their findings with regards to statistical power and effect size measures. The statistical programs are available and accessible, yet researchers seem to ignore the APA's (Wilkinson, 1999) requests to report both statistical power and effect size measures. The current study discovered that about 28% of the studies reported effect size indices while even fewer reported calculating statistical power. Of the studies that did report effect size indices, the vast majority did not interpret them. It seems that the current status of both statistical power and effect size indices is that they are under utilized and not understood.

Future/limitations

Because the field is not evolving to incorporate statistical power in to their pre-study practices and reporting and interpreting effect size measures in their post experiment process, it seems that statistical power surveys and effect size surveys may need to adapt. Using the current methods, statistical power surveys do not seem to impact the field (Sedlmeier & Gigerenzer, 1989). A broad power survey which exams a variety of research topics seems to diffuse its impact. Future statistical power surveys may become oriented to specific topic areas instead of incorporating a spectrum of topics. By researching the statistical power in certain areas (i.e. Post Traumatic Stress

Disorder), psychology researchers may be able to determine which areas are in need of more statistical power.

Statistical power surveys will also have to adapt in terms of incorporating more advanced statistical analysis techniques. While advanced statistical analysis techniques are becoming more prevalent in the literature, only preliminary statistical power methods have been developed for these techniques. In order to complete a statistical power survey on these methods, the researcher must not only make assumptions about effect sizes, but also other variables that influence statistical power. For example, in logistic regression, the researcher would have to make estimates for small, medium, and large for the odds ratio (currently there are no published studies which establish values for this) and the r-squared between the independent variables. This results in not 3 estimated of statistical power as in traditional power surveys but a matrix of possible outcomes. The researcher could estimate 3 levels for the odds ratios (small, medium, and large) and 3 levels for the r-squared (small, medium, and large) resulting in 9 possible values for power.

Whereas statistical power techniques will have to become more advanced, effect size indices will have to become more accessible. Statistical analysis computer programs provide some effect size indices commonly (i.e. odds ratios for logistic regression), but do display other effect size indices by default (i.e. eta-squared). It seems that to increase the use of effect size measures, computer programs will have to report effect size indices by default.

However, generating the effect size indices is only half of the problem. As researchers have stated previously, when effect size indices are being reported, they are rarely being interpreted (Thompson, 1997a, 1997b, 1999, 2002). Journal editors can require that researchers report the indices, but it seems nearly impossible to dictate that they interpret them. Institutions of higher education need to start addressing this issue in the introduction to statistics classes, conferences need to provide seminars training researchers in this skill, and articles need to be written explaining the steps in understanding, interpreting, and reporting effect size indices.

Statistical power and effect size indices have evolved from their initial inception. As the field of psychology continues to progress, these techniques need to continue to advance and the field needs to increase their awareness and understanding of these techniques. These techniques complement the statistical messages that we are already providing in the literature. They provide the punctuation to our statistical sentences and increase other investigators' understanding of our research.

## References

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

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

Neyman, J. & Pearson, E. S. (1928a). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika, 20A, 175-240.

Neyman, J. & Pearson, E. S. (1928b). On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. Biometrika, 20A, 263-294.

Neyman, J. & Pearson, E. S. (1933a). On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London, Series A, 231, 289-337.

Neyman, J. & Pearson, E. S. (1933b). The testing of statistical hypotheses in relation to probabilities a priori. Proceedings of the Cambridge Philosophical Society, 29, 492-510.

Sedlmeier, P. & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? Psychological Bulletin, 105, 309-316.

Thompson, B. (1997a). Editorial policies regarding statistical significance tests: Further comments. Educational Researcher, 26, 29-32.

Thompson, B. (1997b). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. Educational & Psychological Measurement, 57, 5-19.

Thompson, B. (1999). Why "encouraging" effect size reporting is not working: The etiology of researcher resistance to changing practices. The Journal of Psychology, 133, 133-140.

Thompson, B. (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? Journal of Counseling and Development, 80, 64-71.

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

Yule, G. U. (1900). On the association of attributes in statistics. Philosophical Transactions of the Royal Society, A, 194, 257-319.

Appendix A:

Potential Articles in Statistical Power Survey and Effect Size Survey

Aarts, H. & Dijksterhuis, A. (2000). Habits as knowledge structures: Automaticity in goal-directed behavior. Journal of Personality & Social Psychology, 78(1) 53-63.

Abrams, D., Marques, J. M., Bown, N., & Henson, M. (2000). Pro-norm and anti-norm deviance within and between groups. Journal of Personality & Social Psychology, 78(5) 906-912.

Addis, M. E. & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. Journal of Consulting & Clinical Psychology, 68(2) 331-339.

Alloy, L. B., Abramson, L. Y., Hogan, M. E., Whitehouse, W. G., Rose, D. T., Robinson, M. S., Kim, R. S., & Lapkin, J. B. (2000). The Temple-Wisconsin Cognitive Vulnerability to Depression Project: Lifetime history of Axis I psychopathology in individuals at high and low cognitive risk for depression. Journal of Abnormal Psychology, 109(3), 403-418

Amir, N., Foa, E. B., & Coles, M. E. (2000). Implicit memory bias for threat-relevant information in individuals with generalized social phobia. Journal of Abnormal Psychology, 109(4) 713-720.

Anderson, C. A. & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. Journal of Personality & Social Psychology, 78(4) 772-790.

Andrews, B., Brewin, C. R., Rose, S., & Kirk, M. (2000). Predicting PTSD symptoms in victims of violent crime: The role of shame, anger, and childhood abuse. Journal of Abnormal Psychology, 109(1) 69-73.

Andrews, J. A., Foster, S. L., Capaldi, D., & Hops, H. (2000). Adolescent and family predictors of physical aggression, communication, and satisfaction in young adult couples: A prospective analysis. Journal of Consulting & Clinical Psychology, 68(2), 195-208.

Antoni, M. H., Cruess, D. G., Cruess, S., Lutgendorf, S., Kumar, M., Ironson, G., Klimas, N., Fletcher, M. A., & Schneiderman, N. (2000). Cognitive-behavioral stress management intervention effects on anxiety, 24-hr urinary norepinephrine output, and T-cytotoxic/suppressor cells over time among symptomatic HIV-infected gay men. Journal of Consulting & Clinical Psychology, 68(1) 31-45.

Armeli, S., Carney, M. A., Tennen, H., Affleck, G., & O'Neil, T. (2000). Stress and alcohol use: A daily process examination of the stressor-vulnerability model. Journal of Personality & Social Psychology, 78(5) 979-994.

Aron, A., Norman, C. C., Aron, E. N., McKenna, C., & Heyman, R. E. (2000). Couples' shared participation in novel and arousing activities and experienced relationship quality. Journal of Personality & Social Psychology, 78(2) 273-284.

Bailey, J. M., Dunne, M. P., & Martin, N. G. (2000). Genetic and environmental influences on sexual orientation and its correlates in an Australian twin sample. Journal of Personality & Social Psychology, 78(3) 524-536.

Bailey, J. M., Kirk, K. M., Zhu, G., Dunne, M. P., & Martin, N. G. (2000). Do individual differences in sociosexuality represent genetic or environmentally contingent strategies? Evidence from the Australian twin registry. Journal of Personality & Social Psychology, 78(3) 537-545.

Barber, J. P., Connolly, M. B., Crits-Christoph, P., Gladis, L., & Siqueland, L. (2000). Alliance predicts patients' outcome beyond in-treatment change in symptoms. Journal of Consulting & Clinical Psychology, 68(6) 1027-1032.

Barry, C. T., Frick, P. J., DeShazo, T. M., McCoy, M., Ellis, M., & Loney, B. R. (2000). The importance of callous-unemotional traits for extending the concept of psychopathy to children. Journal of Abnormal Psychology, 109(2) 335-340.

Beckham, J. C., Feldman, M. E., Barefoot, J. C., Fairbank, J. A., Helms, M. J., Haney, T. L., Hertzberg, M. A., Moore, S. D., & Davidson, J. R. T. (2000). Ambulatory cardiovascular activity in Vietnam combat veterans with and without posttraumatic stress disorder. Journal of Consulting & Clinical Psychology, 68(2) 269-276.

Beidel, D. C., Turner, S. M., & Morris, T. L. (2000). Behavioral treatment of childhood social phobia. Journal of Consulting & Clinical Psychology, 68(6) 1072-1080.

Benazon, N. R. (2000). Predicting negative spousal attitudes toward depressed persons: A test of Coyne's Interpersonal Model. Journal of Abnormal Psychology, 109(3) 550-554.

Benotsch, E. G., Brailey, K., Vasterling, J. J., Uddo, M., Constans, J. I., & Sutker, P. B. (2000). War Zone stress, personal and environmental resources, and PTSD symptoms in Gulf War Veterans: A longitudinal perspective. Journal of Abnormal Psychology, 109(2) 205-213.

Bickman, L., Lambert, E. W., Andrade, A. R., & Penaloza, R. V. (2000). The Fort Bragg continuum of care for children and adolescents: Mental health outcomes over 5 years. Journal of Consulting & Clinical Psychology, 68(4) 710-716.

Blanchard, J. J., Gangestad, S. W., Brown, S. A., & Horan, W. P. (2000). Hedonic capacity and schizotypy revisited: A taxometric analysis of social anhedonia. Journal of Abnormal Psychology, 109(1) 87-95.

Boldero, J. & Francis, J. (2000). The relation between self-discrepancies and emotion: The moderating roles of self-guide importance, location relevance, and social self-domain centrality. Journal of Personality & Social Psychology, 78(1), 38-52

Borsari, B. & Carey, K. B. (2000). Effects of a brief motivational intervention with college student drinkers. Journal of Consulting & Clinical Psychology, 68(4) 728-733.

Bowler, D. M., Gardiner, J. M., Grice, S., & Saavalainen, P. (2000). Memory illusions: False recall and recognition in adults with Asperger's syndrome. Journal of Abnormal Psychology, 109(4), 663-672.

Braaten, E. B. & Rosen, L. A. (2000). Self-regulation of affect in attention deficit-hyperactivity disorder (ADHD) and non-ADHD boys: Differences in empathic responding. Journal of Consulting & Clinical Psychology, 68(2) 313-321.

Brandon, T. H., Collins, B. N., Juliano, L. M., & Lazev, A. B. (2000). Preventing relapse among former smokers: A comparison of minimal interventions through telephone and mail. Journal of Consulting & Clinical Psychology, 68(1) 103-113.

Brewin, C. R., Andrews, B., & Valentine, J. D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. Journal of Consulting & Clinical Psychology, 68(5) 748-766.

Brown, G. K., Beck, A. T., Steer, R. A., & Grisham, J. R. (2000). Risk factors for suicide in psychiatric outpatients: A 20-year prospective study. Journal of Consulting & Clinical Psychology, 68(3) 371-377.

Brown, P. D. & O'Leary, K. D. (2000). Therapeutic alliance: Predicting continuance and success in group treatment for spouse abuse. Journal of Consulting & Clinical Psychology, 68(2) 340-345.

Bryant, R. A., Harvey, A. G., Guthrie, R. M., & Moulds, M. L. (2000). A prospective study of psychophysiological arousal, acute stress disorder, and posttraumatic stress disorder. Journal of Abnormal Psychology, 109(2) 341-344.

Budney, A. J., Higgins, S. T., Radonovich, K. J., & Novy, P. L. (2000). Adding voucher-based incentives to coping skills and motivational enhancement improves outcomes during treatment for marijuana dependence. Journal of Consulting & Clinical Psychology, 68(6) 1051-1061.

Burns, D. D. & Spangler, D. L. (2000). Does psychotherapy homework lead to improvements in depression in cognitive-behavioral therapy or does improvement lead to increased homework compliance? Journal of Consulting & Clinical Psychology, 68(1) 46-56.

Butler, R. (2000). Making judgments about ability: The role of implicit theories of ability in moderating inferences from temporal and social comparison information. Journal of Personality & Social Psychology, 78(5) 965-978.

Calhoun, P. S., Sampson, W. S., Bosworth, H. B., Feldman, M. E., Kirby, A. C., Hertzberg, M. A., Wampler, T. P., Tate-Williams, F., Moore, S. D., & Beckham, J. C. (2000). Drug use and validity of substance use self-reports in veterans seeking help for posttraumatic stress disorder. Journal of Consulting & Clinical Psychology, 68(5) 923-927.

Cano, A. & O'Leary, K. D. (2000). Infidelity and separations precipitate major depressive episodes and symptoms of nonspecific depression and anxiety. Journal of Consulting & Clinical Psychology, 68(5) 774-781.

Carbonari, J. P. & DiClemente, C. C. (2000). Using transtheoretical model profiles to differentiate levels of alcohol abstinence success. Journal of Consulting & Clinical Psychology, 68(5) 810-817.

Carlson, C. L. & Tamm, L. (2000). Responsiveness of children with attention deficit-hyperactivity disorder to reward and response cost: Differential impact on performance and motivation. Journal of Consulting & Clinical Psychology, 68(1) 73-83.

Carney, M. A., Armeli, S., Tennen, H., Affleck, G., & O'Neil, T. P. (2000). Positive and negative daily events, perceived stress, and alcohol use: A diary study. Journal of Consulting & Clinical Psychology, 68(5) 788-798.

Carson, M. A., Paulus, L. A., Lasko, N. B., Metzger, L. J., Wolfe, J., Orr, S. P., & Pitman, R. K. (2000). Psychophysiologic assessment of posttraumatic stress disorder in Vietnam nurse veterans who witnessed injury or death. Journal of Consulting & Clinical Psychology, 68(5) 890-897.

Carver, C. S., Meyer, B., & Antoni, M. H. (2000). Responsiveness to threats and incentives, expectancy of recurrence, and distress and disengagement: Moderator effects in women with early stage breast cancer. Journal of Consulting & Clinical Psychology, 68(6) 965-975.

Case, L. & Smith, T. B. (2000). Ethnic representation in a sample of the literature of applied psychology. Journal of Consulting & Clinical Psychology, 68(6) 1107-1110.

Caspi, A. (2000). The child is father of the man: Personality continuities from childhood to adulthood. Journal of Personality & Social Psychology, 78(1) 158-172.

Caughlin, J. P., Huston, T. L., & Houts, R. M. (2000). How does personality matter in marriage? An examination of trait anxiety, interpersonal negativity, and marital satisfaction. Journal of Personality & Social Psychology, 78(2) 326-336.

Celio, A. A., Winzelberg, A. J., Wilfley, D. E., Eppstein-Herald, D., Springer, E. A., Dev, P., & Taylor, C. B. (2000). Reducing risk factors for eating disorders: Comparison of an Internet- and a classroom-delivered psychoeducational program. Journal of Consulting & Clinical Psychology, 68(4) 650-657.

Chiu, C. y., Morris, M. W., Hong, Y. y., & Menon, T. (2000). Motivated cultural cognition: The impact of implicit cultural theories on dispositional attribution varies as a function of need for closure. Journal of Personality & Social Psychology, 78(2) 247-259.

Cochran, S. D., Keenan, C., Schober, C., & Mays, V. M. (2000). Estimates of alcohol use and clinical treatment needs among homosexually active men and women in the U.S. population. Journal of Consulting & Clinical Psychology, 68(6) 1062-1071.

Collins, N. L. & Feeney, B. C. (2000). A safe haven: An attachment theory perspective on support seeking and caregiving in intimate relationships. Journal of Personality & Social Psychology, 78(6) 1053-1073.

Conn, R. & Posey, T. B. (2000). Dichotic listening in college students who report auditory hallucinations. Journal of Abnormal Psychology, 109(3) 546-549.

Connor-Smith, J. K., Compas, B. E., Wadsworth, M. E., Thomsen, A. H., & Saltzman, H. (2000). Responses to stress in adolescence: Measurement of coping and involuntary stress responses. Journal of Consulting & Clinical Psychology, 68(6) 976-992.

Cook, W. L. (2000). Understanding attachment security in family context. Journal of Personality & Social Psychology, 78(2) 285-294.

Corning, A. F. (2000). Assessing perceived social inequity: A relative deprivation framework. Journal of Personality & Social Psychology, 78(3) 463-477.

Coyne, J. C., Benazon, N. R., Gaba, C. G., Calzone, K., & Weber, B. L. (2000). Distress and psychiatric morbidity among women from high-risk breast and ovarian cancer families. Journal of Consulting & Clinical Psychology, 68(5) 864-874.

Cross, S. E., Bacon, P. L., & Morris, M. L. (2000). The relational-interdependent self-construal and relationships. Journal of Personality & Social Psychology, 78(4) 191-808.

Cumsille, P. E., Sayer, A. G., & Graham, J. W. (2000). Perceived exposure to peer and adult drinking as predictors of growth in positive alcohol expectancies during adolescence. Journal of Consulting & Clinical Psychology, 68(3) 531-536.

Cunningham, C. E., Boyle, M., Offord, D., Racine, Y., Hundert, J., Secord, M., & McDonald, J. (2000). Tri-ministry study: Correlates of school-based parenting course utilization. Journal of Consulting & Clinical Psychology, 68(5) 928-933.

Currie, S. R., Wilson, K. G., Pontefract, A. J., & deLaplante, L. (2000). Cognitive-behavioral treatment of insomnia secondary to chronic pain. Journal of Consulting & Clinical Psychology, 68(3) 407-416.

Cutting, L. P. & Docherty, N. M. (2000). Schizophrenia outpatients' perceptions of their parents: Is expressed emotion a factor? Journal of Abnormal Psychology, 109(2) 266-272.

Daley, S. E., Burge, D., & Hammen, C. (2000). Borderline personality disorder symptoms as predictors of 4-year romantic relationship dysfunction in young women: Addresing issues of specificity. Journal of Abnormal Psychology, 109(3), 451-460.

Daley, S. E., Hammen, C., & Rao, U. (2000). Predictors of first onset and recurrence of major depression in young women during the 5 years following high school graduation. Journal of Abnormal Psychology, 109(3), 525-533.

Dar, R., Rish, S., Hermesh, H., Taub, M., & Fux, M. (2000). Realism of confidence in obsessive-compulsive checkers. Journal of Abnormal Psychology, 109(4) 673-678.

Davis, P. J. & Gibson, M. G. (2000). Recognition of posed and genuine facial expressions of emotion in paranoid and nonparanoid schizophrenia. Journal of Abnormal Psychology, 109(3) 445-450.

Davison, G. C. (2000). Stepped care: Doing more with less? Journal of Consulting & Clinical Psychology, 68(4) 580-585.

De Dreu, C. K. W., Weingart, L. R., & Kwon, S. (2000). Influence of social motives on integrative negotiation: A meta-analytic review and test of two theories. Journal of Personality & Social Psychology, 78(5) 889-905.

de Jong, P. J. & Merckelbach, H. (2000). Phobia-relevant illusory correlations: The role of phobic responsivity. Journal of Abnormal Psychology, 109(4) 597-601.

Deldin, P. J., Keller, J., Gergen, J. A., & Miller, G. A. (2000). Right-posterior face processing anomaly in depression. Journal of Abnormal Psychology, 109(1) 116-121.

DeSteno, D., Petty, R. E., Wegener, D. T., & Rucker, D. D. (2000). Beyond valence in the perception of likelihood: The role of emotion specificity. Journal of Personality & Social Psychology, 78(3) 397-416.

Dolinski, D. (2000). On inferring one's beliefs from one's attempt and consequences for subsequent compliance. Journal of Personality & Social Psychology, 78(2) 260-272.

Donnellan, M. B., Ge, X., & Wenk, E. (2000). Cognitive abilities in adolescent-limited and life-course-persistent criminal offenders. Journal of Abnormal Psychology, 109(3), 396-402.

Doucet, C. & Stelmack, R. M. (2000). An event-related potential analysis of extraversion and individual differences in cognitive processing speed and response execution. Journal of Personality & Social Psychology, 78(5) 956-964.

Dougall, A. L., Herberman, H. B., Delahanty, D. L., Inslicht, S. S., & Baum, A. (2000). Similarity of prior trauma exposure as a determinant of chronic stress responding to an airline disaster. Journal of Consulting & Clinical Psychology, 68(2) 290-295.

Doyle, A. E., Biederman, J., Seidman, L. J., Weber, W., & Faraone, S. V. (2000). Diagnostic efficiency of neuropsychological test scores for discriminating boys with and without attention deficit-hyperactivity disorder. Journal of Consulting & Clinical Psychology, 68(3) 477-488.

Dunford, F. W. (2000). The San Diego Navy Experiment: An assessment of interventions for men who assault their wives. Journal of Consulting & Clinical Psychology, 68(3) 468-476.

Durbin, C. E., Klein, D. N., & Schwartz, J. E. (2000). Predicting the 2-sup-1/-sub-2-year outcome of dysthymic disorder: The roles of childhood adversity and family history of psychopathology. Journal of Consulting & Clinical Psychology, 68(1) 57-63.

Eddy, J. M. & Chamberlain, P. (2000). Family managment and deviant peer association as mediators of the impact of treatment condition on youth antisocial behavior. Journal of Consulting & Clinical Psychology, 68(5) 857-863.

Edinger, J. D., Fins, A. I., Glenn, D. M., Sullivan, R. J. J., Bastian, L. A., Marsh, G. R., Dailey, D., Hope, T. V., Young, M., Shaw, E., & Vasilas, D. (2000). Insomnia and the eye of the beholder: Are there clinical markers of objective sleep disturbances among adults with and without insomnia complaints? Journal of Consulting & Clinical Psychology, 68(4) 586-593.

Ehlers, A., Maercker, A., & Boos, A. (2000). Posttraumatic stress disorder following political imprisonment: The role of mental defeat, alienation, and perceived permanent change. Journal of Abnormal Psychology, 109(1) 45-55.

Eisenberg, N., Fabes, R. A., Guthrie, I. K., & Reiser, M. (2000). Dispositional emotionality and regulation: Their role in predicting quality of social functioning. Journal of Personality & Social Psychology, 78(1) 136-157.

Epstein, L. H., Paluch, R. A., Gordy, C. C., Saelens, B. E., & Ernst, M. M. (2000). Problem solving in the treatment of childhood obesity. Journal of Consulting & Clinical Psychology, 68(4) 717-721.

Evans, G. W., Wells, N. M., Chan, H. Y. E., & Saltzman, H. (2000). Housing quality and mental health. Journal of Consulting & Clinical Psychology, 68(3) 526-530.

Fals-Stewart, W., O'Farrell, T. J., Freitas, T. T., McFarlin, S. K., & Rutigliano, P. (2000). The Timeline Followback reports of psychoactive substance use by drug-abusing patients: Psychometric properties. Journal of Consulting & Clinical Psychology, 68(1) 134-144.

Faraone, S. V., Biederman, J., Feighner, J. A., & Monuteaux, M. C. (2000). Assessing symptoms of attention deficit hyperactivity disorder in children and adults: Which is more valid? Journal of Consulting & Clinical Psychology, 68(5) 830-842.

Fazio, R. H., Ledbetter, J. E., & Towles-Schwen, T. (2000). On the costs of accessible attitudes: Detecting that the attitude object has changed. Journal of Personality & Social Psychology, 78(2) 197-210.

Finn, P. R., Sharkansky, E. J., Brandt, K. M., & Turcotte, N. (2000). The effects of familial risk, personality, and expectancies on alcohol use and abuse. Journal of Abnormal Psychology, 109(1) 122-133.

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. Journal of Personality & Social Psychology, 78(2) 350-365.

Franklin, M. E., Abramowitz, J. S., Kozak, M. J., Levitt, J. T., & Foa, E. B. (2000). Effectiveness of exposure and ritual prevention for obsessive-compulsive disorder: Randomized compared with nonrandomized samples. Journal of Consulting & Clinical Psychology, 68(4) 594-602.

Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. Journal of Personality & Social Psychology, 78(6) 1135-1149.

Galinsky, A. D. & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. Journal of Personality & Social Psychology, 78(4) 708-724.

Gangestad, S. W., Bailey, J. M., & Martin, N. G. (2000). Taxometric analyses of sexual orientation and gender identity. Journal of Personality & Social Psychology, 78(6) 1109-1121.

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. Journal of Personality & Social Psychology, 78(2) 211-222.

Gleaves, D. H., Lowe, M. R., Snow, A. C., Green, B. A., & Murphy-Eberenz, K. P. (2000). Continuity and discontinuity models of bulimia nervosa: A taxometric investigation. Journal of Abnormal Psychology, 109(1), 56-68.

Gold, J. M., Rehkemper, G., Binks, S. W. I., Carpenter, C. J., Fleming, K., Goldberg, T. E., & Weinberger, D. R. (2000). Learning and forgetting in schizophrenia. Journal of Abnormal Psychology, 109(3) 534-538.

Goldstein, A. J., de Beurs, E., Chambless, D. L., & Wilson, K. A. (2000). EMDR for panic disorder with agoraphobia: Comparison with waiting list and credible attention-placebo control conditions. Journal of Consulting & Clinical Psychology, 68(6) 947-956.

Greenberger, E., Chen, C., Tally, S. R., & Dong, Q. (2000). Family, peer, and individual correlates of depressive symptomatology among U.S. and Chinese adolescents. Journal of Consulting & Clinical Psychology, 68(2) 209-219.

Greeno, C. G., Wing, R. R., & Shiffman, S. (2000). Binge antecedents in obese women with and without binge eating disorder. Journal of Consulting & Clinical Psychology, 68(1) 95-102.

Grisso, T., Davis, J., Vesselinov, R., Appelbaum, P. S., & Monahan, J. (2000). Violent thoughts and violent behavior following hospitalization for mental disorder. Journal of Consulting & Clinical Psychology, 68(3) 388-398.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. Journal of Consulting & Clinical Psychology, 68(1) 155-165.

Grych, J. H., Jouriles, E. N., Swank, P. R., McDonald, R., & Norwood, W. D. (2000). Patterns of adjustment among children of battered women. Journal of Consulting & Clinical Psychology, 68(1) 84-94.

Haaga, D. A. F. (2000). Introduction to the special section on stepped care models in psychotherapy. Journal of Consulting & Clinical Psychology, 68(4) 547-548.

Hagerty, M. R. (2000). Social comparisons of income in one's community: Evidence from national surveys of income and happiness. Journal of Personality & Social Psychology, 78(4) 764-771.

Hammen, C., Gitlin, M., & Altshuler, L. (2000). Predictors of work adjustment in bipolar I patients: A naturalistic longitudinal follow-up. Journal of Consulting & Clinical Psychology, 68(2) 220-225.

Hammen, C., Henry, R., & Daley, S. E. (2000). Depression and sensitization to stressors among young women as a function of childhood adversity. Journal of Consulting & Clinical Psychology, 68(5) 782-787.

Harris, C. R. (2000). Psychophysiological responses to imagined infidelity: The specific innate modular view of jealousy reconsidered. Journal of Personality & Social Psychology, 78(6) 1082-1091.

Haselton, M. G. & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. Journal of Personality & Social Psychology, 78(1) 81-91.

Hassin, R. & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. Journal of Personality & Social Psychology, 78(5) 837-852.

Hatsukami, D. K., Grillo, M., Boyle, R., Allen, S., Jensen, J., Bliss, R., & Brown, S. (2000). Treatment of spit tobacco users with transdermal nicotine system and mint snuff. Journal of Consulting & Clinical Psychology, 68(2) 241-249.

Heatherton, T. F. & Vohs, K. D. (2000). Interpersonal evaluations following threats to self: Role of self-esteem. Journal of Personality & Social Psychology, 78(4) 725-736.

Higgins, S. T., Wong, C. J., Badger, G. J., Ogden, D. E. H., & Dantona, R. L. (2000). Contingent reinforcement increases cocaine abstinence during outpatient treatment and 1 year of follow-up. Journal of Consulting & Clinical Psychology, 68(1) 64-72.

Hilliard, R. B., Henry, W. P., & Strupp, H. H. (2000). An interpersonal model of psychotherapy: Linking patient and therapist developmental history, therapeutic process, and types of outcome. Journal of Consulting & Clinical Psychology, 68(1), 125-133.

Hirsch, C. R. & Mathews, A. (2000). Impaired positive inferential bias in social phobia. Journal of Abnormal Psychology.Vol 109(4) 705-712.

Hlastala, S. A., Frank, E., Kowalski, J., Sherrill, J. T., Tu, X. M., Anderson, B., & Kupfer, D. J. (2000). Stressful life events, bipolar disorder, and the "kindling model". Journal of Abnormal Psychology, 109(4), 777-786.

Hoffman, K. B., Cole, D. A., Martin, J. M., Tram, J., & Seroczynski, A. D. (2000). Are the discrepancies between self- and others' appraisals of competence predictive or reflective of depressive symptoms in children and adolescents: A longitudinal study, part II. Journal of Abnormal Psychology, 109(4) 651-662.

Holahan, C. J., Moos, R. H., Holahan, C. K., & Cronkite, R. C. (2000). Long-term posttreatment functioning among patients with unipolar depression: An integrative model. Journal of Consulting & Clinical Psychology, 68(2) 226-232.

Holtzworth-Munroe, A., Meehan, J. C., Herron, K., Rehman, U., & Stuart, G. L. (2000). Testing the Holtzworth-Munroe and Stuart (1994) batterer typology. Journal of Consulting & Clinical Psychology, 68(6) 1000-1019.

Hooley, J. M. & Hiller, J. B. (2000). Personality and expressed emotion. Journal of Abnormal Psychology, 109(1) 40-44.

Huey, S. J. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. Journal of Consulting & Clinical Psychology, 68(3) 451-467.

Ingram, R. E. & Ritter, J. (2000). Vulnerability to depression: Cognitive reactivity and parental bonding in high-risk individuals. Journal of Abnormal Psychology, 109(4) 588-596.

Jackson, K. M., Sher, K. J., & Wood, P. K. (2000). Prospective analysis of comorbidity: Tobacco and alcohol use disorders. Journal of Abnormal Psychology, 109(4) 679-694.

Jacobson, N. S., Christensen, A., Prince, S. E., Cordova, J., & Eldridge, K. (2000). Integrative behavioral couple therapy: An acceptance-based, promising new treatment for couple discord. Journal of Consulting & Clinical Psychology, 68(2) 351-355.

Ji, L. J., Peng, K., & Nisbett, R. E. (2000). Culture, control, and perception of relationships in the environment. Journal of Personality & Social Psychology, 78(5) 943-955.

Johnson, S. L. & Jacob, T. (2000). Sequential interactions in the marital communication of depressed men and women. Journal of Consulting & Clinical Psychology, 68(1) 4-12.

Johnson, S. L., Sandrow, D., Meyer, B., Winters, R., Miller, I., Solomon, D., & Keitner, G. (2000). Increases in manic symptoms after life events involving goal attainment. Journal of Abnormal Psychology, 109(4) 721-727.

Joiner, T. E. J. & Rudd, M. D. (2000). Intensity and duration of suicidal crisis vary as a function of previous suicide attempts and negative life events. Journal of Consulting & Clinical Psychology, 68(5) 909-916.

Jorm, A. F., Christensen, H., Henderson, A. S., Jacomb, P. A., Korten, A. E., & Rodgers, B. (2000). Predicting anxiety and depression from personality: Is there a synergistic effect of neuroticism and extraversion? Journal of Abnormal Psychology, 109(1) 145-149.

Karney, B. R. & Bradbury, T. N. (2000). Attributions in marriage: State or trait? A growth curve analysis. Journal of Personality & Social Psychology, 78(2) 295-309.

Kassel, J. D. & Unrod, M. (2000). Smoking, anxiety, and attention: Support for the role of nicotine in attentionally mediated anxiolysis. Journal of Abnormal Psychology, 109(1) 161-166.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. Journal of Personality & Social Psychology, 78(5) 871-888.

Kazantzis, N. (2000). Power to detect homework effects in psychotherapy outcome research. Journal of Consulting & Clinical Psychology, 68(1) 166-170.

Keller, J., Nitschke, J. B., Bhargava, T., Deldin, P. J., Gergen, J. A., Miller, G. A., & Heller, W. (2000). Neuropsychological differentiation of depression and anxiety. Journal of Abnormal Psychology, 109(1) 3-10.

Kendall, P. C. & Sheldrick, R. C. (2000). Normative data for normative comparisons. Journal of Consulting & Clinical Psychology, 68(5) 767-773.

Kentgen, L. M., Tenke, C. E., Pine, D. S., Fong, R., Klein, R. G., & Bruder, G. E. (2000). Electroencephalographic asymmetries in adolescents with major depression: Influence of comorbidity with anxiety disorders. Journal of Abnormal Psychology, 109(4) 797-802.

Kerns, J. G. & Berenbaum, H. (2000). Aberrant semantic and affective processing in people at risk for psychosis. Journal of Abnormal Psychology, 109(4) 728-732.

Killen, J. D., Fortmann, S. P., Schatzberg, A. F., Hayward, C., Sussman, L., Rothman, M., Strausberg, L., & Varady, A. (2000). Nicotine patch and paroxetine for smoking cessation. Journal of Consulting & Clinical Psychology, 68(5) 883-889.

Kilpatrick, D. G., Acierno, R., Saunders, B., Resnick, H. S., Best, C. L., & Schnurr, P. P. (2000). Risk factors for adolescent substance abuse and dependence: Data from a national sample. Journal of Consulting & Clinical Psychology, 68(1) 19-30.

King, D. W., King, L. A., Erickson, D. J., Huang, M. T., Sharkansky, E. J., & Wolfe, J. (2000). Posttraumatic stress disorder and retrospectively reported stressor exposure: A longitudinal prediction model. Journal of Abnormal Psychology, 109(4) 624-633.

Klump, K. L., McGue, M., & Iacono, W. G. (2000). Age differences in genetic and environmental influences on eating attitudes and behaviors in preadolescent and adolescent female twins. Journal of Abnormal Psychology, 109(2) 239-251.

Knight, R. A., Manoach, D. S., Elliott, D. S., & Hershenson, M. (2000). Perceptual organization in schizophrenia: The processing of symmetrical configurations. Journal of Abnormal Psychology, 109(4) 575-587.

Kolko, D. J., Brent, D. A., Baugher, M., Bridge, J., & Birmaher, B. (2000). Cognitive and family therapies for adolescent depression: Treatment specificity, mediation, and moderation. Journal of Consulting & Clinical Psychology, 68(4) 603-614.

Korfine, L. & Hooley, J. M. (2000). Directed forgetting of emotional stimuli in borderline personality disorder. Journal of Abnormal Psychology, 109(2), 214-221.

Kremen, W. S., Seidman, L. J., Faraone, S. V., Toomey, R., & Tsuang, M. T. (2000). The paradox of normal neuropsychological function in schizophrenia. Journal of Abnormal Psychology, 109(4), 743-752.

Kwapil, T. R., Miller, M. B., Zinser, M. C., Chapman, L. J., Chapman, J., & Eckblad, M. (2000). A longitudinal study of high scorers on the Hypomanic Personality Scale. Journal of Abnormal Psychology, 109(2) 222-226.

Ladouceur, R., Dugas, M. J., Freeston, M. H., Leger, E., Gagnon, F., & Thibodeau, N. (2000). Efficacy of a cognitive-behavioral treatment for generalized anxiety disorder: Evaluation in a controlled clinical trial. Journal of Consulting & Clinical Psychology, 68(6) 957-964.

Ladouceur, R., Freeston, M. H., Rheaume, J., Dugas, M. J., Gagnon, F., Thibodeau, N., & Fournier, S. (2000). Strategies used with intrusive thoughts: A comparison of OCD patients with anxious and community controls. Journal of Abnormal Psychology, 109(2), 179-187.

Lahey, B. B., Schwab-Stone, M., Goodman, S. H., Waldman, I. D., Canino, G., Rathouz, P. J., Miller, T. L., Dennis, K. D., Bird, H., & Jensen, P. S. (2000). Age and gender differences in oppositional behavior and conduct problems: A cross-sectional household study of middle childhood and adolescence. Journal of Abnormal Psychology, 109(3) 488-503.

Langenbucher, J., Martin, C. S., Labouvie, E., Sanjuan, P. M., Bavly, L., & Pollock, N. K. (2000). Toward the DSM-V: The Withdrawal-Gate Model versus the DSM-IV in the diagnosis of alcohol abuse and dependence. Journal of Consulting & Clinical Psychology, 68(5), 799-809.

Lara, M. E., Klein, D. N., & Kasch, K. L. (2000). Psychosocial predictors of the short-term course and outcome of major depression: A longitudinal study of a nonclinical sample with recent-onset episodes. Journal of Abnormal Psychology, 109(4) 644-650.

Latimer, W. W., Newcomb, M., Winters, K. C., & Stinchfield, R. D. (2000). Adolescent substance abuse treatment outcome: The role of substance abuse problem severity, psychosocial, and treatment factors. Journal of Consulting & Clinical Psychology, 68(4) 684-696.

Lee, A. Y., Aaker, J. L., & Gardner, W. L. (2000). The pleasures and pains of distinct self-construals: The role of interdependence in regulatory focus. Journal of Personality & Social Psychology, 78(6) 1122-1134.

Lepore, S. J., Ragan, J. D., & Jones, S. (2000). Talking facilitates cognitive-emotional processes of adaptation to an acute stressor. Journal of Personality & Social Psychology, 78(3) 499-508.

Levenston, G. K., Patrick, C. J., Bradley, M. M., & Lang, P. J. (2000). The psychopath as observer: Emotion and attention in picture processing. Journal of Abnormal Psychology, 109(3), 373-385.

Lewinsohn, P. M., Solomon, A., Seeley, J. R., & Zeiss, A. (2000). Clinical implications of "subthreshold" depressive symptoms. Journal of Abnormal Psychology, 109(2) 345-351.

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. Journal of Personality & Social Psychology, 78(2) 223-246.

Litz, B. T., Orsillo, S. M., Kaloupek, D., & Weathers, F. (2000). Emotional processing in posttraumatic stress disorder. Journal of Abnormal Psychology, 109(1) 26-39.

Lorch, E. P., Milich, R., Sanchez, R. P., van den Broek, P., Baer, S., Hooks, K., Hartung, C., & Welsh, R. (2000). Comprehension of televised stories in boys with attention deficit/hyperactivity disorder and nonreferred boys. Journal of Abnormal Psychology, 109(2) 321-330.

Lynam, D. R., Caspi, A., Moffit, T. E., Wikstroem, P. O., Loeber, R., & Novak, S. (2000). The interaction between impulsivity and neighborhood context on offending: The effects of impulsivity are stronger in poorer neighborhoods. Journal of Abnormal Psychology, 109(4) 563-574.

MacDonald, T. K., Fong, G. T., Zanna, M. P., & Martineau, A. M. (2000). Alcohol myopia and condom use: Can alcohol intoxication be associated with more prudent behavior? Journal of Personality & Social Psychology, 78(4), 605-619.

Mackinger, H. F., Pachinger, M. M., Leibetseder, M. M., & Fartacek, R. R. (2000). Autobiographical memories in women remitted from major depression. Journal of Abnormal Psychology, 109(2) 331-334.

Marcus-Newhall, A., Pedersen, W. C., Carlson, M., & Miller, N. (2000). Displaced aggression is alive and well: A meta-analytic review. Journal of Personality & Social Psychology, 78(4), 670-689.

Marsh, H. W., Kong, C. K., & Hau, K. T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. Journal of Personality & Social Psychology, 78(2) 337-349.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. Journal of Consulting & Clinical Psychology, 68(3) 438-450.

Mathews, A. & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. Journal of Abnormal Psychology, 109(4) 602-615.

McCarroll, J. E., Thayer, L. E., Liu, X., Newby, J. H., Norwood, A. E., Fullerton, C. S., & Ursano, R. J. (2000). Spouse abuse recidivism in the U.S. Army by gender and military status. Journal of Consulting & Clinical Psychology, 68(3) 521-525.

McCrae, R. R., Costa, P. T. J., Ostendorf, F., Angleitner, A., Hrebickova, M., Avia, M. D., Sanz, J., Sanchez-Bernardos, M. L., Kusdil, M. E., Woodfield, R., Saunders, P. R., & Smith, P. B. (2000). Nature over nurture: Temperament, personality, and life span development. Journal of Personality & Social Psychology, 78(1), 173-186.

McCullough, J. P. J., Klien, D. N., Keller, M. B., Holzer, C. E. I., Davis, S. M., Kornstein, S. G., Howland, R. H., Thase, M. E., & Harrison, W. M. (2000). Comparison of DSM-III-R chronic major depression and major depression superimposed on dysthymia (double depression): Validity of the distinction. Journal of Abnormal Psychology, 109(3) 419-427.

McDermott, P. A., Alterman, A. I., Cacciola, J. S., Rutherford, M. J., Newman, J. P., & Mulholland, E. M. (2000). Generality of Psychopathy Checklist--Revised factors over prisoners and substance-dependent patients. Journal of Consulting & Clinical Psychology, 68(1) 181-186.

McNally, R. J., Clancy, S. A., Schacter, D. L., & Pitman, R. K. (2000). Cognitive processing of trauma cues in adults reporting repressed, recovered, or continuous memories of childhood sexual abuse. Journal of Abnormal Psychology, 109(3) 355-359.

McNally, R. J., Clancy, S. A., Schacter, D. L., & Pitman, R. K. (2000). Personality profiles, dissociations, and absorption in women reporting repressed, recovered, or continuous memories of childhood sexual abuse. Journal of Consulting & Clinical Psychology, 68(6) 1033-1037.

McNiel, D. E., Lam, J. N., & Binder, R. L. (2000). Relevance of interrater agreement to violence risk assessment. Journal of Consulting & Clinical Psychology, 68(6) 1111-1115.

McQuaid, J. R., Monroe, S. M., Roberts, J. E., Kupfer, D. J., & Frank, E. (2000). A comparison of two life stress assessment approaches: Prospective prediction of treatment outcome in recurrent depression. Journal of Abnormal Psychology, 109(4) 787-791.

Mesman, J. & Koot, H. M. (2000). Common and specific correlates of preadolescent internalizing and externalizing psychopathology. Journal of Abnormal Psychology, 109(3) 428-437.

Meston, C. M. & Heiman, J. R. (2000). Sexual abuse and sexual function: An examination of sexually relevant cognitive processes. Journal of Consulting & Clinical Psychology, 68(3) 399-406.

Mikulincer, M., Birnbaum, G., Woddis, D., & Nachmias, O. (2000). Stress and accessibility of proximity-related thoughts: Exploring the normative and intraindividual components of attachment theory. Journal of Personality & Social Psychology, 78(3) 509-523.

Miller, I. W., McDermut, W., Gordon, K. C., Keitner, G. I., Ryan, C. E., & Norman, W. (2000). Personality and family functioning in families of depressed patients. Journal of Abnormal Psychology, 109(3) 539-545.

Miller, J. N. & Ozonoff, S. (2000). The external validity of Asperger disorder: Lack of evidence from the domain of neuropsychology. Journal of Abnormal Psychology, 109(2) 227-238.

Mogg, K., Millar, N., & Bradley, B. P. (2000). Biases in eye movements to threatening facial expressions in generalized anxiety disorder and depressive disorder. Journal of Abnormal Psychology, 109(4) 695-704.

Mohr, D. C., Likosky, W., Bertagnolli, A., Goodkin, D. E., Van Der Wende, J., Dwyer, P., & Dick, L. P. (2000). Telephone-administered cognitive-behavioral therapy for the treatment of depressive symptoms in multiple sclerosis. Journal of Consulting & Clinical Psychology, 68(2) 356-361.

Morey, L. C. & Zanarini, M. C. (2000). Borderline personality: Traits and disorder. Journal of Abnormal Psychology, 109(4) 733-737.

Murphy, P. E., Ciarrocchi, J. W., Piedmont, R. L., Cheston, S., Peyrot, M., & Fitchett, G. (2000). The relation of religious belief and practices, depression, and hopelessness in persons with clinical depression. Journal of Consulting & Clinical Psychology.Special Issue:, 68(6) 1102-1106.

Murray, S. L., Holmes, J. G., & Griffin, D. W. (2000). Self-esteem and the quest for felt security: How perceived regard regulates attachment processes. Journal of Personality & Social Psychology, 78(3) 478-498.

Mussell, M. P., Mitchell, J. E., Crosby, R. D., Fulkerson, J. A., Hoberman, H. M., & Romano, J. L. (2000). Commitment to treatment goals in prediction of group cognitive-behavioral therapy treatment outcome for women with bulimia nervosa. Journal of Consulting & Clinical Psychology, 68(3) 432-437.

Mussweiler, T. & Strack, F. (2000). The use of category and exemplar knowledge in the solution of anchoring tasks. Journal of Personality & Social Psychology, 78(6) 1038-1052.

Newman, M. G. (2000). Recommendations for a cost-offset model of psychotherapy allocation using generalized anxiety disorder as an example. Journal of Consulting & Clinical Psychology, 68(4) 549-555.

Nezlek, J. B., Hampton, C. P., & Shean, G. D. (2000). Clinical depression and day-to-day social interaction in a community sample. Journal of Abnormal Psychology, 109(1) 11-19.

Nishith, P., Mechanic, M. B., & Resick, P. A. (2000). Prior interpersonal trauma: The contribution to current PTSD symptoms in female rape victims. Journal of Abnormal Psychology, 109(1) 20-25.

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. Journal of Abnormal Psychology, 109(3) 504-511.

Oishi, S., Wyer, R. S. J., & Colcombe, S. J. (2000). Cultural variation in the use of current life satisfaction to predict the future. Journal of Personality & Social Psychology, 78(3) 434-445.

Oldehinkel, A. J., Ormel, J., & Neeleman, J. (2000). Predictors of time to remission from depression in primary care patients: Do some people benefit more from positive life change than others? Journal of Abnormal Psychology, 109(2) 299-307.

Orr, S. P., Metzger, L. J., Lasko, N. B., Macklin, M. L., Peri, T., & Pitman, R. K. (2000). De novo conditioning in trauma-exposed individuals with and without posttraumatic stress disorder. Journal of Abnormal Psychology, 109(2) 290-298.

Otto, M. W., Pollack, M. H., & Maki, K. M. (2000). Empirically supported treatments for panic disorder: Costs, benefits, and stepped care. Journal of Consulting & Clinical Psychology, 68(4) 556-563.

Palfai, T. P., Monti, P. M., Ostafin, B., & Hutchison, K. (2000). Effects of nicotine deprivation on alcohol-related information processing and drinking behavior. Journal of Abnormal Psychology, 109(1) 96-105.

Park, J. & Banaji, M. R. (2000). Mood and heuristics: The influence of happy and sad states on sensitivity and bias in stereotyping. Journal of Personality & Social Psychology, 78(6) 1005-1023.

Pedersen, W. C., Gonzales, C., & Miller, N. (2000). The moderating effect of trivial triggering provocation on displaced aggression. Journal of Personality & Social Psychology, 78(5) 913-927.

Penn, D. L., Combs, D. R., Ritchie, M., Francis, J., Cassisi, J., Morris, S., & Townsend, M. (2000). Emotion recognition in schizophrenia: Further investigation of generalized versus specific deficit models. Journal of Abnormal Psychology, 109(3) 512-516.

Perkins, K. A., Levine, M., Marcus, M., Shiffman, S., D'Amico, D., Miller, A., Keins, A., Ashcom, J., & Broge, M. (2000). Tobacco withdrawal in women and menstrual cycle phase. Journal of Consulting & Clinical Psychology, 68(1) 176-180.

Peters, E. R., Pickering, A. D., Kent, A., Glasper, A., Irani, M., David, A. S., Day, S., & Hemsley, D. R. (2000). The relationship between cognitive inhibition and psychotic symptoms. Journal of Abnormal Psychology, 109(3) 386-395.

Petry, N. M., Martin, B., Cooney, J. L., & Kranzler, H. R. (2000). Give them prizes and they will come: Contingency management for treatment of alcohol dependence. Journal of Consulting & Clinical Psychology, 68(2) 250-257.

Piasecki, T. M., Niaura, R., Shadel, W. G., Abrams, D., Goldstein, M., Fiore, M. C., & Baker, T. B. (2000). Smoking withdrawal dynamics in unaided quitters. Journal of Abnormal Psychology, 109(1) 74-86.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. Journal of Personality & Social Psychology, 78(3) 582-593.

Pontari, B. A. & Schlenker, B. R. (2000). The influence of cognitive load on self-presentation: Can cognitive busyness help as well as harm social performance? Journal of Personality & Social Psychology, 78(6), 1092-1108.

Rassovsky, Y., Kushner, M. G., Schwarze, N. J., & Wangensteen, O. D. (2000). Psychological and physiological predictors of response to carbon dioxide challenge in individuals with panic disorder. Journal of Abnormal Psychology, 109(4) 616-623.

Reynolds, K. J., Turner, J. C., & Haslam, S. A. (2000). When are we better than them and they worse than us? A closer look at social discrimination in positive and negative domains. Journal of Personality & Social Psychology, 78(1) 64-80.

Richards, J. M., Beal, W. E., Seagal, J. D., & Pennebaker, J. W. (2000). Effects of disclosure of traumatic events on illness behavior among psychiatric prison inmates. Journal of Abnormal Psychology, 109(1) 156-160.

Roberts, L. J., Neal, D. J., Kivlahan, D. R., Baer, J. S., & Marlatt, G. A. (2000). Individual drinking changes following a brief intervention among college students: Clinical significance in an indicated preventive context. Journal of Consulting & Clinical Psychology, 68(3) 500-505.

Rohsenow, D. J., Monti, P. M., Hutchison, K. E., Swift, R. M., Colby, S. M., & Kaplan, G. B. (2000). Naltrexone's effects on reactivity to alcohol cues among alcoholic men. Journal of Abnormal Psychology, 109(4) 738-742.

Rohsenow, D. J., Monti, P. M., Martin, R. A., Michalec, E., & Abrams, D. B. (2000). Brief coping skills treatment for cocaine abuse: 12-month substance use outcomes. Journal of Consulting & Clinical Psychology, 68(3) 515-520.

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: Origins and outcomes. Journal of Personality & Social Psychology, 78(1) 122-135.

Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., & Price, L. (2000). A controlled study of virtual reality exposure therapy for the fear of flying. Journal of Consulting & Clinical Psychology, 68(6) 1020-1026.

Rotherman-Borus, M. J., Piacentini, J., Cantwell, C., Belin, T. R., & Song, J. (2000). The 18-month impact of an emergency room intervention for adolescent female suicide attempters. Journal of Consulting & Clinical Psychology, 68(6) 1081-1093.

Rotton, J. & Cohn, E. G. (2000). Violence is a curvilinear function of temperature in Dallas: A replication. Journal of Personality & Social Psychology, 78(6) 1074-1081.

Rueter, M. A., Chao, W., & Conger, R. D. (2000). The effect of systematic variation in retrospective conduct disorder reports on antisocial personality disorder diagnoses. Journal of Consulting & Clinical Psychology, 68(2), 307-312.

Ruscio, J. & Ruscio, A. M. (2000). Informing the continuity controversy: A taxometric analysis of depression. Journal of Abnormal Psychology, 109(3) 473-487.

Ruscio, J. (2000). The role of complex thought in clinical prediction: Social accountability and the need for cognition. Journal of Consulting & Clinical Psychology, 68(1) 145-154.

Rusting, C. L. & DeHart, T. (2000). Retrieving positive memories to regulate negative mood: Consequences for mood-congruent memory. Journal of Personality & Social Psychology, 78(4), 737-752.

Rychtarik, R. G., Connors, G. J., Whitney, R. B., McGillicuddy, N. B., Fitterling, J. M., & Wirtz, P. W. (2000). Treatment settings for persons with alcoholism: Evidence for matching clients to inpatient versus outpatient care. Journal of Consulting & Clinical Psychology, 68(2) 277-289.

Samoilov, A., Goldfried, M. R., & Shapiro, D. A. (2000). Coding system of Therapeutic Focus on Action and Insight. Journal of Consulting & Clinical Psychology, 68(3) 513-514.

Sanders, M. R., Markie-Dadds, C., Tully, L. A., & Bor, W. (2000). The Triple P-Positive Parenting Program: A comparison of enhanced, standard, and self-directed behavioral family intervention for parents of children with early onset conduct problems. Journal of Consulting & Clinical Psychology, 68(4) 624-640.

Saucier, G. (2000). Isms and the structure of social attitudes. Journal of Personality & Social Psychology, 78(2) 366-385.

Schimel, J., Pyszczynski, T., Greenberg, J., O'Mahen, H., & Arndt, J. (2000). Running from the shadow: Psychological distancing from others to deny characteristics people fear in themselves. Journal of Personality & Social Psychology, 78(3) 446-462.

Schmidt, N. B. & Woolaway-Bickel, K. (2000). The effects of treatment compliance on outcome in cognitive-behavioral therapy for panic disorder: Quality versus quantity. Journal of Consulting & Clinical Psychology, 68(1) 13-18.

Schmidt, N. B., Storey, J., Greenberg, B. D., Santiago, H. T., Li, Q., & Murphy, D. L. (2000). Evaluating gene * psychological risk factor effects in the pathogenesis of anxiety: A new model approach. Journal of Abnormal Psychology, 109(2) 308-320.

Schmidt, N. B., Woolaway-Bickel, K., Trakowski, J., Santiago, H., Storey, J., Koselka, M., & Cook, J. (2000). Dismantling cognitive-behavioral treatment for panic disorder: Questioning the utility of breathing retraining. Journal of Consulting & Clinical Psychology, 68(3) 417-424.

Schnurr, P. P., Ford, J. D., Friedman, M. J., Green, B. L., Dain, B. J., & Sengupta, A. (2000). Predictors and outcomes of posttraumatic stress disorder in World War II veterans exposed to mustard gas. Journal of Consulting & Clinical Psychology, 68(2), 258-268.

Schretlen, D., Jayaram, G., Maki, P., Park, K., Abebe, S., & DiCarlo, M. (2000). Demographic, clinical, and neurocognitive correlates of everyday functional impairment in severe mental illness. Journal of Abnormal Psychology, 109(1) 134-138.

Schulz-Hardt, S., Frey, D., Luethgens, C., & Moscovici, S. (2000). Biased information search in group decision making. Journal of Personality & Social Psychology, 78(4) 655-669.

Schwartz, D. & Proctor, L. J. (2000). Community violence exposure and children's social adjustment in the school peer group: The mediating roles of emotion regulation and social cognition. Journal of Consulting & Clinical Psychology, 68(4) 670-683.

Seidman, L. J., Biederman, J., Monuteaux, M. C., Weber, W., & Faraone, S. V. (2000). Neuropsychological functioning in nonreferred siblings of children with attention deficit/hyperactivity disorder. Journal of Abnormal Psychology, 109(2) 252-265.

Semple, S. J., Patterson, T. L., & Grant, I. (2000). The sexual negotiation behavior of HIV-positive gay and bisexual men. Journal of Consulting & Clinical Psychology, 68(5) 934-937.

Sharkansky, E. J., King, D. W., King, L. A., Wolfe, J., Erickson, D. J., & Stokes, L. R. (2000). Coping with Gulf War combat stress: Mediating and moderating effects. Journal of Abnormal Psychology, 109(2) 188-197.

Sheppard, L. C. & Teasdale, J. D. (2000). Dysfunctional thinking in major depressive disorder: A deficit in metacognitive monitoring? Journal of Abnormal Psychology, 109(4) 768-776.

Shepperd, J. A., Findley-Klein, C., Kwavnick, K. D., Walker, D., & Perez, S. (2000). Bracing for loss. Journal of Personality & Social Psychology, 78(4) 620-634.

Sher, K. J., Bartholow, B. D., & Wood, M. D. (2000). Personality and substance use disorders: A prospective study. Journal of Consulting & Clinical Psychology, 68(5) 818-829.

Shiner, R. L. (2000). Linking childhood personality with adaptation: Evidence for continuity and change across time into late adolescence. Journal of Personality & Social Psychology, 78(2) 310-325.

Sigmon, S. T., Dorhofer, D. M., Rohan, K. J., Hotovy, L. A., Boulard, N. E., & Fink, C. M. (2000). Psychophysiological, somatic, and affective changes across the menstrual cycle in women with panic disorder. Journal of Consulting & Clinical Psychology, 68(3) 425-431.

Smith, B. H., Pelham, W. E. J., Gnagy, E., Molina, B., & Evans, S. (2000). The reliability, validity, and unique contributions of self-report by adolescents receiving treatment for attention-deficit/hyperactivity disorder. Journal of Consulting & Clinical Psychology, 68(3), 489-499.

Sobell, M. B. & Sobell, L. C. (2000). Stepped care as a heuristic approach to the treatment of alcohol problems. Journal of Consulting & Clinical Psychology, 68(4) 573-579.

Stanton, A. L., Danoff-Burg, S., Cameron, C. L., Bishop, M., Collins, C. A., Kirk, S. B., Sworowski, L. A., & Twillman, R. (2000). Emotionally expressive coping predicts psychological and physical adjustment to breast cancer. Journal of Consulting & Clinical Psychology, 68(5) 875-882.

Stanton, A. L., Kirk, S. B., Cameron, C. L., & Danoff-Burg, S. (2000). Coping through emotional approach: Scale construction and validation. Journal of Personality & Social Psychology, 78(6) 1150-1169.

Stapel, D. A. & Koomen, W. (2000). How far do we go beyond the information given? The impact of knowledge activation on interpretation and inference. Journal of Personality & Social Psychology, 78(1) 19-37.

Stein, K. D., Goldman, M. S., & Del Boca, F. K. (2000). The influence of alcohol expectancy priming and mood manipulation on subsequent alcohol consumption. Journal of Abnormal Psychology, 109(1), 106-115.

Stephens, R. S., Roffman, R. A., & Curtin, L. (2000). Comparison of extended versus brief treatments for marijuana use. Journal of Consulting & Clinical Psychology, 68(5) 898-908.

Stevens, A. A., Donegan, N. H., Anderson, M., Goldman-Rakic, P. S., & Wexler, B. E. (2000). Verbal processing deficits in schizophrenia. Journal of Abnormal Psychology, 109(3) 461-471.

Stice, E., Hayward, C., Cameron, R. P., Killen, J. D., & Taylor, C. B. (2000). Body-image and eating disturbances predict onset of depression among female adolescents: A longitudinal study. Journal of Abnormal Psychology, 109(3) 438-444.

Stoolmiller, M., Eddy, J. M., & Reid, J. B. (2000). Detecting and describing preventive intervention effects in a universal school-based randomized trial targeting delinquent and violent behavior. Journal of Consulting & Clinical Psychology, 68(2) 296-306.

Stroem, L., Pettersson, R., & Andersson, G. (2000). A controlled trial of self-help treatment of recurrent headache conducted via the Internet. Journal of Consulting & Clinical Psychology, 68(4) 722-727.

Stuart, G. L., Treat, T. A., & Wade, W. A. (2000). Effectiveness of an empirically based treatment for panic disorder delivered in a service clinic setting: 1-year follow-up. Journal of Consulting & Clinical Psychology, 68(3) 506-512.

Swendsen, J. D., Tennen, H., Carney, M. A., Affleck, G., Willard, A., & Hromi, A. (2000). Mood and alcohol consumption: An experience sampling test of the self-medication hypothesis. Journal of Abnormal Psychology, 109(2) 198-204.

Tarrier, N., Kinney, C., McCarthy, E., Humphreys, L., Wittkowski, A., & Morris, J. (2000). Two-year follow-up of cognitive-behavioral therapy and supportive counseling in the treatment of persistent symptoms in chronic schizophrenia. Journal of Consulting & Clinical Psychology, 68(5) 917-922.

Taylor, J., Iacono, W. G., & McGue, M. (2000). Evidence for a genetic etiology of early-onset delinquency. Journal of Abnormal Psychology, 109(4) 634-643.

Teasdale, J. D., Segal, Z. V., Williams, J. M., Ridgeway, V. A., Soulsby, J. M., & Lau, M. A. (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. Journal of Consulting & Clinical Psychology, 68(4) 615-623.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. Journal of Personality & Social Psychology, 78(5) 853-870.

Thom, A., Sartory, G., & Joehren, P. (2000). Comparison between one-session psychological treatment and benzodiazepine in dental phobia. Journal of Consulting & Clinical Psychology, 68(3) 378-387.

Tiffany, S. T., Cox, L. S., & Elash, C. A. (2000). Effects of transdermal nicotine patches on abstinence-induced and cue-elicited craving in cigarette smokers. Journal of Consulting & Clinical Psychology, 68(2) 233-240.

Titone, D., Levy, D. L., & Holzman, P. S. (2000). Contextual insensitivity in schizophrenic language processing: Evidence from lexical ambiguity. Journal of Abnormal Psychology, 109(4) 761-767.

Tram, J. M. & Cole, D. A. (2000). Self-perceived competence and the relation between life events and depressive symptoms in adolescence: Mediator or moderator? Journal of Abnormal Psychology, 109(4) 753-760.

Trierweiler, S. J., Neighbors, H. W., Munday, C., Thompson, E. E., Binion, V. J., & Gomez, J. P. (2000). Clinician attributions associated with the diagnosis of schizophrenia in African American and non-African American patients. Journal of Consulting & Clinical Psychology, 68(1) 171-175.

Urada, D. I. & Miller, N. (2000). The impact of positive mood and category importance on crossed categorization effects. Journal of Personality & Social Psychology, 78(3) 417-433.

Van den Oord, E. J. C. G., Boomsma, D. I., & Verhulst, F. C. (2000). A study of genetic and environmental effects on the co-occurrence of problem behaviors in three-year-old-twins. Journal of Abnormal Psychology, 109(3) 360-372.

van Dijk, E. & Wilke, H. (2000). Decision-induced focusing in social dilemmas: Give-some, keep-some, take-some, and leave-some dilemmas. Journal of Personality & Social Psychology, 78(1) 92-104.

Ventura, J., Nuechterlein, K. H., Subotnik, K. L., Hardesty, J. P., & Mintz, J. (2000). Life events can trigger depressive exacerbation in the early course of schizophrenia. Journal of Abnormal Psychology, 109(1) 139-144.

Vorauer, J. D., Hunter, A. J., Main, K. J., & Roy, S. A. (2000). Meta-stereotype activation: Evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction. Journal of Personality & Social Psychology, 78(4) 690-707.

Waldo, T. G. & Merritt, R. D. (2000). Family proneness, dissociaton, and DSM-IV Axis II symptomatology. Journal of Abnormal Psychology, 109(3) 555-558.

Waltz, J., Babcock, J. C., Jacobson, N. S., & Gottman, J. M. (2000). Testing a typology of batterers. Journal of Consulting & Clinical Psychology, 68(4) 658-669.

Ward, A. & Mann, T. (2000). Don't mind if I do: Disinhibited eating under cognitive load. Journal of Personality & Social Psychology, 78(4) 753-763.

Watkins, P. C., Martin, C. K., & Stern, L. D. (2000). Unconscious memory bias in depression: Perceptual and conceptual processes. Journal of Abnormal Psychology, 109(2) 282-289.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. Journal of Personality & Social Psychology, 78(3) 546-558.

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. Journal of Personality & Social Psychology, 78(1) 105-119.

Weisman, A. G., Nuechterlein, K. H., Goldstein, M. J., & Snyder, K. S. (2000). Controllability perceptions and reactions to symptoms of schizophrenia: A within-family comparison of relatives with high and low expressed emotion. Journal of Abnormal Psychology, 109(1) 167-171.

Weiss, B., Catron, T., & Harris, V. (2000). A 2-year follow-up of the effectiveness of traditional child psychotherapy. Journal of Consulting & Clinical Psychology, 68(6) 1094-1101.

Wendel, J. S., Miklowitz, D. J., Richards, J. A., & George, E. L. (2000). Expressed emotion and attributions in the relatives of bipolar patients: An analysis of problem-solving interactions. Journal of Abnormal Psychology, 109(4), 792-796.

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. Journal of Personality & Social Psychology, 78(6) 1024-1037.

Whisman, M. A., Sheldon, C. T., & Goering, P. (2000). Psychiatric disorders and dissatisfaction with social relationships: Does type of relationship matter? Journal of Abnormal Psychology, 109(4) 803-808.

Wigboldus, D. H. J., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. Journal of Personality & Social Psychology, 78(1) 5-18.

Wilfley, D. E., Friedman, M. A., Dounchis, J. Z., Stein, R. I., Welch, R. R., & Ball, S. A. (2000). Comorbid psychopathology in binge eating disorder: Relation to eating disorder severity at baseline and following treatment. Journal of Consulting & Clinical Psychology, 68(4) 641-649.

Williams, J. M., Teasdale, J. D., Segal, Z. V., & Soulsby, J. (2000). Mindfulness-based cognitive therapy reduces overgeneral autobiographical memory in formerly depressed patients. Journal of Abnormal Psychology, 109(1) 150-155.

Wilson, A. E. & Ross, M. (2000). The frequency of temporal-self and social comparisons in people's personal appraisals. Journal of Personality & Social Psychology, 78(5) 928-942.

Wilson, G. T., Vitousek, K. M., & Loeb, K. L. (2000). Stepped care treatment for eating disorders. Journal of Consulting & Clinical Psychology, 68(4) 564-572.

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. Journal of Personality & Social Psychology, 78(5) 821-836.

Winzelberg, A. J., Eppstein, D., Eldredge, K. L., Wilfley, D., Dasmahapatra, R., Dev, P., & Taylor, C. B. (2000). Effectiveness of an Internet-based program for reducing risk factors for eating disorders. Journal of Consulting & Clinical Psychology, 68(2) 346-350.

Wolchik, S. A., West, S. G., Sandler, I. N., Tein, J. Y., Coatsworth, D., Lengua, L., Weiss, L., Anderson, E. R., Greene, S. M., & Griffin, W. A. (2000). An experimental evaluation of theory-based mother and mother-child programs for children of divorce. Journal of Consulting & Clinical Psychology, 68(5) 843-856.

Wolsko, C., Park, B., Judd, C. M., & Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. Journal of Personality & Social Psychology, 78(4) 635-654.

Ybarra, G. J., Passman, R. H., & Eisenberg, C. S. (2000). The presence of security blankets or mothers (or both) affects distress during pediatric examinations. Journal of Consulting & Clinical Psychology, 68(2), 322-330.

Yen, S., Robins, C. J., & Lin, N. (2000). A cross-cultural comparison of depressive symptom manifestation: China and the United States. Journal of Consulting & Clinical Psychology, 68(6) 993-999.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. Journal of Consulting & Clinical Psychology, 68(6) 1038-1050.

Yurek, D., Farrar, W., & Andersen, B. L. (2000). Breast cancer surgery: Comparing surgical groups and determining individual differences in postoperative sexuality and body change stress. Journal of Consulting & Clinical Psychology, 68(4) 697-709.

Zadra, A. & Donderi, D. C. (2000). Nightmares and bad dreams: Their prevalence and relationship to well-being. Journal of Abnormal Psychology, 109(2) 273-281.

Zoellner, L. A., Foa, E. B., Brigidi, B. D., & Przeworski, A. (2000). Are trauma victims susceptible to "false memories?". Journal of Abnormal Psychology, 109(3) 517-524.

Zurbriggen, E. L. (2000). Social motives and cognitive power-sex associations: Predictors of aggressive sexual behavior. Journal of Personality & Social Psychology, 78(3), 559-581.

Zuroff, D. C., Blatt, S. J., Sotsky, S. M., Krupnick, J. L., Martin, D. J., Sanislow, C. A. I., & Simmens, S. (2000). Relation of therapeutic alliance and perfectionism to outcome in brief outpatient treatment of depression. Journal of Consulting & Clinical Psychology, 68(1) 114-124.

## Bibliography

Acklin, M. W., McDowell, C. J., & Orndoff, S. (1992). Statistical power and the Rorschach: 1975-1991. Journal of Personality Assessment, 59, 366-379.

Aguinis, H., Pierce, C. A., & Stone-Romero, E. F. (1994). Estimating the power to detect dichotomous moderators with moderated multiple regression. Educational & Psychological Measurement, 54, 690-692.

Alf, E., & Abrahams, N. M. (1968). Relationship between percent overlap and measure of correlation. Educational and Psychological Measurement, 28, 779-792.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the Chi-square test. Journal of the American Statistical Association, 33, 526-536.

Bezeau, S. & Graves, R. (2001). Statistical power and effect sizes of clinical neuropsychology research. Journal of Clinical & Experimental Neuropsychology, 23, 399-406.

Brewer, J. K. & Owen, P. W. (1973). A note on the power of statistical tests in the Journal of Educational Measurement. Journal of Educational Measurement, 10, 71-74.

Brewer, J. K. (1972). On the power of statistical tests in the American Educational Research Journal. American Educational Research Journal, 9, 391-401.

Brown, D. J. (1975). Mirror, mirror... down with the linear model. American Educational Research Journal, 12, 491-505.

Brown, J. & Hale, M. S. (1992). The power of statistical studies in consultation-liason psychiatry. Psychosomatics, 33, 437-443.

Buchner, A., Erdfelder, E., & Faul, F. (1997). How to Use G*Power. Retrieved on April 23, 2002 from http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/how_to_use_gpower.html.

Cady-Webster, K., Hevey, C., Huang, J., & Rossi, J. (2000). Not all power corrupts: Statistical power of research published in the psychology of women quarterly. Paper presented at the Annual Convention of the Eastern Psychological Association, Baltimore, MD.

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, revisited. Journal of Experimental Education, 61, 287-292.

Chase, L. J. & Chase, R. B. (1976). A statistical power analysis of applied psychological research. Journal of Applied Psychology, 61, 234-237.

Chase, L. J. & Tucker, R. K. (1975). A power-analytic examination of contemporary communication research. Speech Monographs, 42, 29-41.

Chase, L. J. & Tucker, R. K. (1976). Statistical power: Derivation, development, and data-analytic implications. Psychological Record, 26, 473-486.

Chow, S. L. (1998). Precis of statistical significance: Rationale, validity, and utility. Behavioral and Brain Sciences, 21, 169-239.

Christensen, J. E. & Christensen, C. E. (1977). Statistical power analysis of health, physical education, and recreation research. Research Quarterly, 48, 204-208.

Clark-Carter, D. (1997). The account taken of statistical power in research published in the British Journal of Psychology. British Journal of Psychology, 88, 71-83.

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

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

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

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

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

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

Cooper, H., & Findley, M. (1982). Expected effect sizes: Estimates for statistical power analysis in social psychology. Personality and Social Psychology Bulletin, 8, 168-173.

Cowles, M. (1989). Statistics in psychology: A historical perspective. Hillsdale, NJ: Lawrence Erlbaum.

Cowles, M. P. & Davis, C. (1982). On the origins of the .05 level of statistical significance. American Psychologist, 37, 553-558.

Craig, J. R., Eison, C. L., & Metze, L. P. (1976). Significance testes and their interpretation: An example utilizing published research and $\omega^2$. Bulletin of the Psychonomic Society, 7, 280-282.

Cramer, H. (1946). Mathematical methods of statistics. Princeton, NJ: Princeton University Press.

D'Amico, E. J., Neilands, T. B., & Zambarano, R. (2001). Power analysis for multivariate and repeated measures designs: A flexible approach using the SPSS MANOVA procedure. Behavior Research Methods, Instruments & Computers, 33, 479-484.

Daly, J. A. & Hexamer, A. (1983). Statistical power in research in English education. Research in the Teaching of English, 17, 157-164.

Dayton, C. M., Schaffer, W. D., & Rogers, B. G. (1973). On appropriate uses and interpretations of power analysis: A comment. American Educational Research Journal, 10, 231-234.

Dilullo, L. K. (1998). A post hoc power analysis of inferential research examining the relationship between mathematics anxiety and mathematics performance. (Doctoral Dissertation, Auburn University, 1998). Dissertation Abstracts International, 58, 2574.

Dunnette, M. D. (1966). Personnel selection and placement. Belmont, Ca: Wadsworth.

Dwyer, J. H. (1974). Analysis of variance and the magnitude of effects. Psychological Bulletin, 81, 731-737.

Dyer, F. J. (1980). Application of power analysis concepts to test reliability research. Educational and Psychological Measurement, 40, 301 - 306.

Elmore, P. B., & Rotou, O. (2001). A primer on basic effect size concepts. Presented at the Annual meeting of the American Educational Research Association, Seattle, WA.

Fagely, N. S., & McKinney, I. J. (1983). Reviewer bias for statistical significant results: A re-examination. Journal of Counseling Psychology, 30, 298-300.

Fagley, N. S. (1985). Applied statistical power analysis and the interpretation of nonsignificant results by research consumers. Journal of Counseling Psychology, 32, 391-396.

Ferris, C. D., Grubbs, F. E., & Weaver, C. L. (1946). Operating characteristics for the common statistical tests of significance. Annals of Mathematical Statistician, 17, 178-197.

Ferron, J. & Sentovich, C. (2002). Statistical power of randomization tests used with multiple-baseline designs. Journal of Experimental Education, 70, 165-178.

Fisher, R. A. (1932). Statistical methods for research workers (4th ed.). London: Oliver & Boyd.

Frick, R. W. (1999). Defending the status quo. Theory & Psychology, 9, 183-189.

Gillet, R. (1994). Post Hoc power analysis. Journal of Applied Psychology, 79, 783-785.

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, 5, 3-8.

Glass, G., & Stanley, J. (1970). Statistical methods in education and psychology. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Haase, R. F. (1976). Power analysis of research in counselor education. Counselor Education and Supervision, 14, 124-132.

Hamblin, R. L. (1971). Mathematical experimentation and sociological theory: A critical analysis. Sociometry, 34, 423-452.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). What if there were no significance tests? Mahwah, NJ: Erlbaum.

Hays, W. L. (1963). Statistics for psychologists. New York: Holt, Rinehart, & Winston.

Hedges, L. V. & Olkin, I. (1985). Statistical models for meta-analysis. New York: Academic Press.

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

Hedges, L.V. & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. Psychological Methods, 6, 203-217.

Hintze, J.L. (2000). PASS 2000 user's guide. Kaysville, Utah: Number Cruncher Statistical Systems.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. Annals of Mathematical Statistics, 23, 169-192.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. The American Statistician, 55, 19-24.

Hogarty, K. Y., & Kromrey, J. D. (2001, April). We've bee reporting some effect sizes: Can we guess what they mean? Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Hogben, L. (1957). Statistical theory: The relationship of probability, credibility, and error. An examination of the contemporary crisis in statistical theory from a behaviorist viewpoint. London: Allen & Unwin.

Huberty, C. J. (1994) A note on interpreting an $R^2$ value. Journal of Educational and Behavioral Statistics, 19, 351-356.

Huberty, C. J. (2001, April) A History of Effect Size Indices. Presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Huberty, C. J. (2002) A history of effect size indices. Educational and Psychological Measurement, 62, 227-240.

Hunter, J. E. (1997). Needed: A ban on the significance test. Psychological Science, 8, 3-7.

Indurkhya, A. & von Eye, A. (2000). The power of tests in Configural Frequency Analysis. Psychologische Beitrage, 42, 301-308.

Jones, B. J. & Brewer, J. K. (1972). An analysis of the power of statistical tests reported in the Research Quarterly. Research Quarterly, 43, 23-30.

Katzer, J. & Sodt, J. (1973). An analysis of the use of statistical testing in communication research. Journal of Communication, 23, 251-265.

Kazantzis, N. (2000). Power to detect homework effects in psychotherapy outcome research. Journal of Consulting & Clinical Psychology, 68, 166-170.

Kazdin, A. E., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. Journal of Consulting & Clinical Psychology, 57, 138-147.

Kelley, T. L. (1920). Measurement of overlapping. Journal of Educational Psychology, 11, 458-461.

Kelley, T. L. (1923). Statistical method. New York: Macmillan.

Kelley, T. L. (1935). An unbiased correlation ratio. Proceedings of the National Academy of Sciences, 21, 554-559.

Kenna, G., & Rossi, J. S. (2002, March) Statistical power in the journal "Experimental and Clinical Psychopharmacology." Poster session presented at the 73rd Eastern Psychological Association convention, Boston, MA.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olenjnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of educational researchers: An anlysis of there ANOVA, MANOVA, and ANCOVA analysis. Review of Educational Research, 68, 350-386.

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

Knapp, T. R. (1998) Comments on the statistical significance testing articles. Research in the Schools, 5, 39-41.

Kosciulek, J. F. (1993). The statistical power of vocational evaluation research. Vocational Evaluation and Work Adjustment Bulletin, 26, 142-145.

Kosciulek, J. F. & Szymanski, E. M. (1993). Statistical power analysis of rehabilitation counseling research. Rehabilitation Counseling Bulletin, 36, 212-220.

Kroll, R. M. & Chase, L. J. (1975). Communication disorders: A power-analytic assessment of recent research. Journal of Communication Disorders, 8, 237-247.

Lehmann, E. L. (1953). The power of rank tests. Annals of Mathematical Statistics, 24, 23-43.

Levenson, R. L. (1980). Statistical power analysis: Implications for researchers, planners, and practitioners in gerontology. Gerontologist, 20, 494-498.

Levin, J. R., & Robinson, D. H. (1999). Further reflections on hypothesis testing and editorial policy for primary research journals. Educational Psychology Review, 11, 143-155.

Linhart, H. (1957). Power functions of tests concerning the product moment correlation coefficient. Journal National Institutional Personnel Research, 7, 51-65.

Lipsey, M. W. (1990) Design Sensitivity Statistical power for Experimental Research. London: Sage Publications, Inc.

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1, 130-149.

Maddock, J. E., & Rossi, J. S. (2001). Statistical power of articles published in three health-psychology related journals. Health Psychology, 20, 76-78.

Maxwell, S. E., Camp, J. C., & Arvey, R. D. (1981). Measures of strength of association: A comparative examination. Journal of Applied Psychology, 66, 525-534.

Mazen, A. M., Graf, L. A., Kellog, C. E., & Hemmasi, M. (1987). Statistical power in contemporary management research. Academy of Management Journal, 30, 1105-1112.

Mazen, A. M., Hemmasi, M., & Lewis, M. F., (1987). Assessment of statistical power in contemporary strategy research. Strategic Management Journal, 8, 403-410.

McGraw, K. O. & Wong S. P. (1992). The descriptive use of absolute differences between pairs of scores with a common mean and variance. Journal of Educational Statistics, 19,103-110.

McNamara, J. F. (1978). Practical significance and statistical models. Educational Administration Quarterly, 14, 48-63.

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

Mitchell, C., & Hartmann, D. P. (1981). A cautionary note on the use of omega-squared to evaluate the effectiveness of behavioral treatments. Behavioral Assessment, 3, 93-100.

Mone, M. A., Mueller, G. C., & Mauland, W. (1996). The perceptions and usage of statistical power in applied psychology and management research. Personnel Psychology, 49, 103-120.

Morrison, D.E., & Henkle, R. E. (Eds.). (1970). The significance test controversy. Chicago: Aldine.

Mosteller, F. & Bush, R. R. (1954). Selected quantitative techniques. In G. Lindzey (Ed.), Handbook of Social Psychology: Theory and Method (Vol. 1, pp. 289-334). Reading, MA: Addison-Wesley.

Murray, L. W., & Dosser, D. A., Jr. (1987). How significant is a significant difference? Problems with the measurement of magnitude of effect. Journal of Counseling Psychology, 34, 68-72.

Neyman, J. & Pearson, E. S. (1928a). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika, 20A, 175-240.

Neyman, J. & Pearson, E. S. (1928b). On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. Biometrika, 20A, 263-294.

Neyman, J. & Pearson, E. S. (1933a). On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London, Series A, 231, 289-337.

Neyman, J. & Pearson, E. S. (1933b). The testing of statistical hypotheses in relation to probabilities a priori. Proceedings of the Cambridge Philosophical Society, 29, 492-510.

O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. Psychological Bulletin, 92, 766-777.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. Contemporary Educational Psychology, 25, 241-286.

Orme, J. G. & Tolman, R. M. (1986). The statistical power of a decade of social work education research. Social Service Review, 60, 620-632.

Orme, J. G., & Combs-Orme, T. D. (1986). Statistical power and Type II errors in social work research. Social Work Research & Abstracts, 22, 3-10.

Ottenbacher, K. (1982). Statistical power of research in occupational therapy. Occupational Therapy Journal of Research, 2, 13-25.

Patnaik, P. B. (1949). The non-central $\chi^2$ and F distributions and their applications. Biometrika, 36, 203-232.

Pearson, K. (1914). On certain errors with regard to multiple correlation occasionally made by those who have not adequately studies this subject. Biometrika, 10, 181-187.

Pearson, K., & Lee, A. (1897). On the distribution of frequency (variation and correlation) of the barometric height of divers stations. Philosophical Transactions of the Royal Society of London, 190, 423-469.

Pennick, J. E., & Brewer, J. K. (1972). The power of statistical tests in science teaching research. Journal of Research in Science Teaching, 9, 377-381.

Peters, C. C., & Van Voorhis, W. R. (1940). Statistical procedures and their mathematical bases. New York: Mc Graw-Hill.

Reed, J. F., & Slaichert, W. (1980). Statistical proof in inconclusive 'negative' trials. Archives of Internal Medicine, 141, 1307-1310.

Richardson, J. T. E. (1996). Measures of effect size. Behavioral Research Methods, Instruments, & Computers, 28, 12-22.

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

Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. Journal of Applied Social Psychology, 9, 395-396.

Rosnow, R. L. & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. Psychological Methods, 1, 331-340.

Rossi, J. S. (1984). Statistical power of psychological research. Unpublished doctoral dissertation, University of Rhode Island.

Rossi, J. S. (1988). ONEWAY: A BASIC program for computing ANOVA from group summary statistics. Behavior Research Methods, Instruments & Computers, 20, 347-348.

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? Journal of Consulting and Clinical Psychology, 58, 646-656.

Rubin, A., & Conway, P. G. (1985). Standards for determining the magnitude of relationships in social work research. Social Work Research & Abstracts, 21, 34-39.

Sanchez, J., Valera, A., Velandrino, A. P. & Marin, F. (1992). Un studio de la potencia estadistica en Anales de Psicologia (1984 - 1991). Anales de Psicologia, 8, 19-32.

Sawyer, A. G. & Ball, A. D. (1981). Statistical power and effect size in marketing research. Marketing Research, 18, 275-290.

Schmidt, F. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. American Psychologist, 47, 1173-1181.

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

Schuster, C. & von Eye, A. (2000). Using Log-Linear Modeling to increase power in two-sample Configural Frequency Analysis. Psychologische Beitrage, 42, 273-284.

Sechrest, L., & Yeaton, W. H. (1982). Magnitudes of experimental effects in social science research. Evaluation Review, 6, 579-600.

Sedlmeier, P. & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? Psychological Bulletin, 105, 309-316.

Sherron, R. H. (1988). Power analysis: The other half of the coin. Community/ Junior College Quarterly, 12, 169-175.

Sindelar, P.T., Allman, C. Monda, L., Vail, C. O., Wilson, C. L. & Schloss, P. J. (1988). The power of hypothesis testing in special education efficacy research. Journal of Special Education, 22, 284-296.

Spreitzer, E. & Chase, L. J. (1974). Statistical power in sociological research: An examination of data-analytic strategies. Unpublished manuscript, Bowling Green State University, Department of Sociology, Bowling Green, Oh.

Stigler, S. M. (1986). The history of statistics. Cambridge, MA: Belknap.

Strube, M. J. (1988). Some comments on the use of magnitudes-of-effects estimates. Journal of Counseling Research, 35, 342-345.

Stuart, A. (1952). The power of two difference-sign tests. Journal of the American Statistical Association, 47, 416-424.

Tang, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. Statistical Research Memoirs, 2, 126-149.

Tatsuoka, M. M. (1970). Discriminant analysis: The study of group differences. Champaign, IL: Institute for Personality and Ability Testing.

Tatsuoka, M. M. (1973). An examination of the statistical properties of a multivariate measure of strength of association. Final Report to U.S. Office of Education on Contract No. OEG-5-72-0027.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25, 26-30.

Thompson, B. (1997a). Editorial policies regarding statistical significance tests: Further comments. Educational Researcher, 26, 29-32.

Thompson, B. (1997b). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. Educational & Psychological Measurement, 57, 5-19.

Thompson, B. (1999). Why "encouraging" effect size reporting is not working: The etiology of researcher resistance to changing practices. The Journal of Psychology, 133, 133-140.

Thompson, B. (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? Journal of Counseling and Development, 80, 64-71.

Thompson, B., & Snyder, P. A. (1997). Statistical significance testing practices in the Journal of Experimental Education. Journal of Experimental Education, 66, 75-83.

Thompson, B., & Snyder, P. A. (1998). Statistical significance and reliability analyses in recent JCD research articles. Journal of Counseling and Development, 76, 436-441.

Tilton, J. W. (1937). The measurement of overlapping. Journal of Eduacational Psychology, 28, 656-662.

Utts, J. (1988). Successful replication versus statistical significance. Journal of Parapsychology, 52, 305-320.

Valera, A., Sanchez, J., Marin, F., & Velandrino, A. P. (1998). Potencia estadistica de la Revista de Psicologia General y Aplicada. Revista de Psicologia General y Aplicada, 51, 233-246.

Whittington, C. J., Podd, J., & Kan, M. M. (2000). Recognition memory impairment in Parkinson's disease: Power and meta-analyses. Neuropsychology, 14, 233-246.

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

Wilks, S. S. (1932). Certain generalizations of the analysis of variance. Biometrika, 39, 471-494.

Winer, B. J. (1962). Statistical principles in experimental design. New York: McGraw-Hill.

Wolfowitz, J. (1949). The power of the classical tests associated with the normal distribution. Annals of Mathematical Statistics, 20, 540-551.

Wooley, T. W. & Dawson, G. O. (1983). A follow-up power analysis of the tests used in Journal of Research in Science Teaching. Journal of Research in Science Teaching, 20, 673-681.

Wooley, T.W. (1983). A comprehensive power-analytic investigation of research in medical education. Journal of Medical Education, 58, 710-715.

Yule, G. U. (1900). On the association of attributes in statistics. Philosophical Transactions of the Royal Society, A, 194, 257-319.