Professional Psychology: Research and Practice

Copy of e-mail Notification z2h2269

Dr. Ferguson: Article 2009-0620 (z2h - 2269) is available for download

_____

Professional Psychology: Research and Practice Published by American Psychological Association

Dear Author,

The page proof of your article (# 2009-0620), which is scheduled for publication in the Professional Psychology: Research and Practice, is now ready for your final review. To access your proof, please refer to this URL:

http://rapidproof.cadmus.com/RapidProof/retrieval/index.jsp

Login: your e-mail address
Password: siq7XCsgU3fx

The site contains one file. You will need to have Adobe Acrobat® Reader software (Version 4.0 or higher) to read it. This free software is available for user downloading at http://www.adobe.com/products/acrobat/readstep.html.

If you have any problems with downloading your article from the Rapid Proof site, please contact rapidprooftech@cadmus.com. Please include your article number (2009-0620) with all correspondence.

This file contains a reprint order form, information regarding subscriptions and special offers, instructions for correcting the PDF proof electronically, and a copy of the page proof for your article. The proof contains 7 pages.

After printing the PDF file, please read the page proof carefully, as it will not be proofread by APA staff, and email your changes to the Journal Production Manager at APAProduction@cadmus.com within 48 hours. Be sure to respond to any queries that appear on the last page of the proof. Proofread the following elements of your article especially carefully:

- Tables
- Equations and mathematical symbols
- Figures (including checking figure and caption placement)
- Non-English characters and symbols

If you have no changes, email the Journal Production Manager at APAProduction@cadmus.com that you have no changes. If you do have changes, use the comments and notes features in Adobe Acrobat (instructions provided in the PDF file to be downloaded) and return the annotated PDF to APAProduction@cadmus.com. If you are unable to use Adobe Acrobat and your changes are minimal, you may summarize them in an email to the Journal Production Manager, clearly indicating the location of each change. If you are unable to use Adobe Acrobat and have changes too extensive to summarize in an e-mail please proceed as follows: Within the continental United States, send a clearly marked proof to the postal address given at the end of this message; outside the continental United States, fax a clearly marked proof to the Journal Production Manager at the fax number given at the end of this message.

To order reprints, please fill out the reprint order form and return it to Cadmus Reprints, Reprint Account Manager, P.O. Box 751903, Charlotte, NC 28275-1903.

For information about the NIH PubMed Central deposit request as it relates to articles published in APA journals, including a PubMed Central deposit request form, please go to http://www.apa.org/journals/authors/pubmed-deposit.html.

This article has not yet been selected to appear in an issue. The Editor makes all decisions concerning publication order.

If you have questions or concerns about the editing of your proof, please contact me at the email address below.

Sincerely,

Chuck Fulkerson
Journal Production Manager
Professional Psychology: Research and Practice
Cadmus Communications
8621 Robert Fulton Drive, Suite 100
Columbia, MD 21046
E-mail: APAProduction@cadmus.com
Phone: 410-691-6995
FAX: 410-684-2792

Adding Comments and Notes to Your PDF

To facilitate electronic transmittal of corrections, we encourage authors to utilize the comments and notes features in Adobe Acrobat. The PDF provided has been "comment- enabled," which allows you to utilize the comments and notes features, even if using only the free Adobe Acrobat reader (see note below regarding acceptable versions). Adobe Acrobat's Help menu provides additional details on the tool. When you open your PDF, the comments/notes/edit tools are clearly shown on the tool bar. For purposes of correcting the PDF proof of your journal article, the important features to know are the following:

- Note tool (yellow text balloon icon): Click on this feature on the tool bar and then click on a point of the PDF where you would like to make a comment. This feature is useful when providing an instruction to your production contact or to verify a question or change that was posed.

- Text edits ("T" with strike-through and carat icon): The option Insert Text at Cursor allows the user to place the cursor at a point in text, which will then provide a pop-up box to add the appropriate text to be added to the proof.

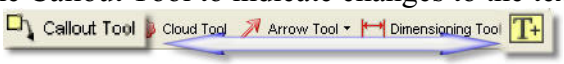
- Highlight text ("T" with yellow highlight strike-through): The option Cross-Out Text Tool allows the user to indicate text that should be deleted from the proof. Click on this feature on the tool bar and highlight the text that is to be deleted.

As with hand-annotated proof corrections, the important points are to communicate changes clearly and thoroughly; to answer all queries and questions; and to provide complete information for us to make the necessary changes to your article so it is ready for publication.

To utilize the comments/notes features on this PDF you will need Adobe Reader version 7 or higher. This program is freely available and can be downloaded from http://www.adobe.com/products/acrobat/readstep2.html.

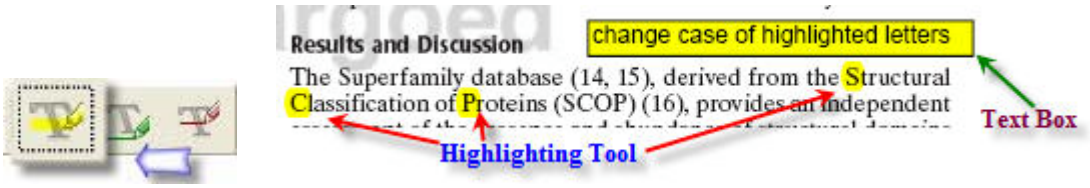
Instructions for E-proofing

- Use Text Boxes and the Callout Tool to indicate changes to the text.

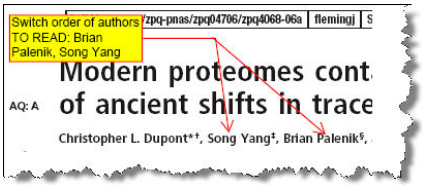


- Use the Strike-Out tool to indicate deletions to the text.

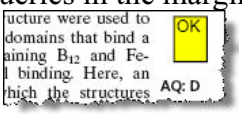
- Use the Highlighting Tool to indicate font problems, bad breaks, and other textual inconsistencies.



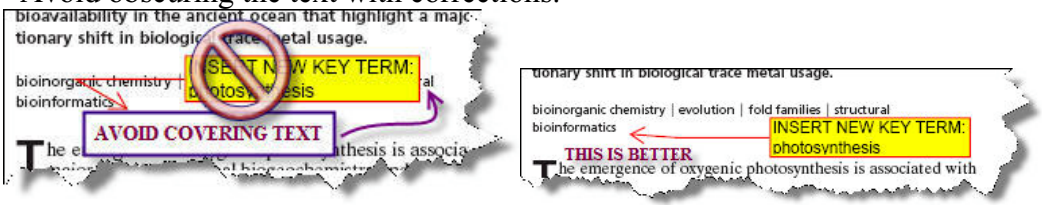
- Clearly indicate where changes need to be made using arrow, lines, and the Call-Out Tool.



- Mark changes and answer queries in the margins and other areas of white space.



- Avoid obscuring the text with corrections.



REPRINTS

Authors have two options for ordering reprints:
Authors who need to use purchase orders may order reprints from the standard *Purchase Order Service.* For a substantially lower price, authors may use the *Prepaid Service.* The rate schedules on page 3 give the prices for each service. All international prices include shipping via WWDS. All domestic shipments will be made via UPS. We request that you do not use Post Office box numbers; provide a full street address if possible. Authors may request expedited shipping – the additional cost will be billed. You are required to pay all duties if they apply.

- **Prepaid Service:** To take advantage of this lower price option, submit your credit card information with your order or enclose a money order, certified check, or personal check. The prices given on page 3 include postage.

- **Purchase Order Service:** Reprint orders that are not prepaid must be accompanied by a purchase order. Cadmus Reprints will bill you later for the cost of the reprints. **Do not send remittance with the reprint order form and purchase order.** Remember that the price you see on page 3 includes postage, so it is the exact amount you will be billed. (Exception: Authors requesting expedited shipping will be billed the additional cost.)

Complete the order form on the next page and return it to Cadmus Reprints (not to APA). Only one order form is provided – include your coauthors' orders on this form or make photocopies of the order form for your coauthors. Check the box for either the prepaid service or the purchase order service. Give explicit instructions for all authors receiving reprints, using the space at the bottom of the order form.

To determine the cost of the reprints, count the number of pages in the printed article and refer to the rate schedules on page 3. For example, if your article is 11 pages long, you want 100 reprints, you live in the United States, and you are using the prepaid service, your total cost would be $118. If your proof includes a page of queries following the text, do not include the query page in your article count.

Send the order form to Cadmus Reprints when you return the page proofs. Reprints will be mailed within two weeks of the publication of the journal. Orders received after the issue goes to press will be processed with the next issue.

Where to Order Reprints

Send your order form with credit card information, money order, certified check, personal check, or purchase order in the amount indicated on the rate schedule to:

Cadmus Reprints
P.O. Box 751903
Charlotte, NC 28275-1903

Phone: (410) 819-3914 FAX: (410) 820-9765

Personal checks must clear before reprints are processed. There is a $30.00 charge for returned checks.

AMERICAN PSYCHOLOGICAL ASSOCIATION

2009 REPRINT ORDER FORM

APA Journal Authors: To order reprints, complete all sections of this form. Please read the instructions on page 1.

Professional Psychology: Research and Practice

3472754

2009-0620

SEND the reprint order and
(1) credit card number, or
(2) money order/certified check, or
(3) approved purchase order, or
(4) check
to:

Cadmus Reprints
P.O. Box 751903
Charlotte, NC 28275-1903

BILLING NAME

ORGANIZATION

ADDRESS (no P.O. Box)

CITY STATE ZIP CODE

ARTICLE TITLE

AUTHOR

DATE

PHONE #

FAX #

E-MAIL

REPRINTS INCLUDE COLOR? ☐ YES ☐ NO

OF TOTAL ARTICLE PAGES

OF REPRINTS

OF COVERS

PAYMENT METHOD: CHECK ONE:

___ CREDIT CARD CARD NUMBER _____
_____ VISA EXPIRATION DATE _____
_____ MASTERCARD SIGNATURE _____
___ MONEY ORDER/CERT. CHECK (make payable to **Cadmus Reprints**)
___ APPROVED PURCHASE ORDER (original PO must be attached)
___ CHECK (shipment delayed until check clears)

COMPLETE SHIPPING LABEL below. No P.O. Boxes.
Include phone number on international shipments. International shipments can be made via AIR for additional charges; please indicate in Special Shipping Instructions if you desire this expedited service.

COMPUTE COST OF ORDER

PRICE (per chart)	$
Add'l for Covers	$
SUBTOTAL	$
PA residents add 6% tax	$
Add'l for Expedited Shipping	$
TOTAL	$

(TYPE OR PRINT MAILING LABEL BELOW)

SHIP TO: Phone No. _____
Name _____
Address _____
_____
City _____ State _____ Zip _____
Expedited Service (enter service required): _____

AMERICAN PSYCHOLOGICAL ASSOCIATION

RATES EFFECTIVE WITH 2009 ISSUES

Black and White Reprint Prices Prepaid						
Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	$56	$75	$84	$114	$145	$175
5-8	$72	$96	$144	$189	$237	$286
9-12	$92	$124	$194	$269	$344	$421
13-16	$105	$153	$248	$344	$441	$536
17-20	$124	$182	$298	$423	$542	$663
21-24	$142	$208	$353	$498	$641	$789
Covers	$66	$73	$97	$121	$148	$173

Color Reprint Prices Prepaid						
Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	$118	$199	$280	$376	$436	$477
5-8	$184	$240	$358	$523	$633	$740
9-12	$203	$266	$413	$607	$798	$944
13-16	$219	$287	$472	$690	$911	$1,129
17-20	$240	$310	$532	$779	$1,023	$1,318
21-24	$263	$334	$595	$866	$1,138	$1,512
Covers	$66	$73	$97	$121	$148	$173

Black and White Reprint Prices Prepaid						
International (includes Canada and Mexico)						
# of Pages	50	100	200	300	400	500
1-4	$82	$98	$120	$170	$215	$265
5-8	$112	$132	$214	$291	$371	$452
9-12	$149	$179	$295	$421	$538	$665
13-16	$178	$225	$381	$539	$698	$855
17-20	$215	$272	$464	$667	$861	$1,057
21-24	$246	$312	$548	$787	$1,022	$1,259
Covers	$93	$99	$142	$195	$245	$296

Color Reprint Prices Prepaid						
International (includes Canada and Mexico)						
# of Pages	50	100	200	300	400	500
1-4	$144	$223	$316	$431	$507	$568
5-8	$224	$276	$429	$624	$766	$906
9-12	$260	$322	$515	$759	$992	$1,188
13-16	$292	$359	$605	$886	$1,169	$1,448
17-20	$332	$401	$698	$1,023	$1,343	$1,711
21-24	$367	$438	$790	$1,154	$1,519	$1,983
Covers	$93	$99	$142	$195	$245	$296

Black and White Reprint Prices Purchase Order						
Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	$61	$81	$91	$124	$157	$190
5-8	$78	$104	$156	$205	$257	$311
9-12	$100	$134	$210	$292	$373	$457
13-16	$114	$166	$269	$373	$478	$581
17-20	$134	$198	$323	$459	$588	$720
21-24	$154	$225	$383	$541	$696	$856
Covers	$71	$79	$105	$131	$160	$188

Color Reprint Prices Purchase Order						
Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	$128	$216	$304	$408	$474	$518
5-8	$200	$261	$389	$567	$687	$803
9-12	$220	$289	$448	$659	$865	$1,024
13-16	$238	$311	$512	$749	$989	$1,224
17-20	$261	$337	$578	$846	$1,110	$1,430
21-24	$285	$362	$645	$939	$1,235	$1,641
Covers	$71	$79	$105	$131	$160	$188

Black and White Reprint Prices Purchase Order						
International (includes Canada and Mexico)						
# of Pages	50	100	200	300	400	500
1-4	$89	$106	$130	$184	$234	$288
5-8	$121	$143	$232	$315	$402	$490
9-12	$162	$195	$320	$457	$584	$721
13-16	$193	$244	$414	$585	$758	$928
17-20	$234	$296	$503	$724	$934	$1,147
21-24	$267	$338	$595	$853	$1,109	$1,366
Covers	$101	$107	$154	$211	$266	$321

Color Prices Purchase Order						
International (includes Canada and Mexico)						
# of Pages	50	100	200	300	400	500
1-4	$156	$242	$343	$468	$550	$616
5-8	$243	$300	$465	$677	$832	$983
9-12	$282	$349	$558	$824	$1,076	$1,289
13-16	$317	$389	$657	$961	$1,268	$1,571
17-20	$360	$435	$758	$1,110	$1,457	$1,857
21-24	$398	$475	$857	$1,252	$1,648	$2,151
Covers	$101	$107	$154	$211	$266	$321

Additional Rates	
Set title page, each	$16.00
Each extra mailing	$32.00
Remake pages, each	$50.00

Subscriptions and Special Offers

In addition to purchasing reprints of their articles, authors may purchase an annual subscription, purchase an individual issue of the journal (at a reduced rate), or request an individual issue at no cost under special "hardship" circumstances.

To place your order, fill out the order form below (including the mailing label) and send the completed form and your check or credit card information to the address listed on the order form.

For information about becoming a member of the American Psychological Association, call the Membership Office at 1-800-374-2721.

2009 APA Journal Subscription Rates

Journal*	Individual Rate	APA Member Rate
American Psychologist	$ 275	$ 12
Behavioral Neuroscience	$ 286	$ 148
Developmental Psychology	$ 249	$ 114
Emotion	$ 103	$ 55
Experimental & Clinical Psychopharm	$ 149	$ 55
Health Psychology	$ 159	$ 65
Jrnl of Abnormal Psychology	$ 159	$ 70
Jrnl of Applied Psychology	$ 249	$ 98
Jrnl of Comparative Psychology	$ 89	$ 55
Jrnl of Consulting & Clinical Psychology	$ 249	$ 118
Jrnl of Counseling Psychology	$ 129	$ 55
Jrnl of Educational Psychology	$ 159	$ 76
JEP: Animal Behavior Processes	$ 89	$ 55
JEP: Applied	$ 89	$ 55
JEP: General	$ 89	$ 55
JEP: Human Perception & Performance	$ 349	$ 157
JEP: Learning, Memory & Cognition	$ 349	$ 157

Journal*	Individual Rate	APA Member Rate
Jrnl of Family Psychology	$ 149	$ 55
Jrnl of Personality & Social Psychology	$ 488	$ 221
Neuropsychology	$ 149	$ 55
Professional Psych.: Research & Practice	$ 129	$ 55
Psychological Assessment	$ 149	$ 55
Psychological Bulletin	$ 249	$ 99
Psychological Methods	$ 89	$ 55
Psychological Review	$ 159	$ 74
Psychology & Aging	$ 159	$ 70
Psychology of Addictive Behaviors	$ 89	$ 73
Psychology, Public Policy & Law	$ 89	$ 55
Rehabilitation Psychology	$ 89	$ 59

*For journal descriptions, see APA's Web site:
http://www.apa.org/journals

CUT ALONG DASHED LINE – RETURN LOWER PORTION TO APA ORDER DEPARTMENT

Instructions: Check the appropriate box, enter journal title and price information, and complete the mailing label in the right column. Enclose a check made out to the **American Psychological Association**, and mail it with the form to the APA Order Department or complete the credit card information below .

☐ Annual Subscription (available on January-December basis only). To subscribe, specify calendar year of the subscription. Refer to the Subscription Rates shown above.

Journal: _____

Calendar year of subscription: _____ Price: _____

Special Offers! If you are an APA journal article author, you may take advantage of two Special Offers. (These offers do not apply to Educational Publishing Foundation journals.)

☐ **Individual Copy.** You may order individual copies of the entire issue in which your article appears. As an author, you receive a special reduced rate of $5 per copy for up to a maximum of 25 copies. No phone requests accepted.

Journal: _____

Vol. no.: _____ Issue no.: _____ Issue month: _____

____ copies @ $5 a copy = $_____ (order amount)
+_____ (handling; see below)
TOTAL enclosed: $_____

Handling Fees		
Order amount:	U.S.	International
< $15	$5	$15
$15 - $60	$6	$16
> $60	Order amount X .10	$20

☐ **Hardship Request.** If you do not have a personal subscription to the journal and you do not have access to an institutional or departmental subscription copy, you may obtain a single copy of the issue in which your article appears at no cost by filing out the information below.

Journal: _____

Vol. no. : _____ Issue no.: _____

Issue month: _____

CREDIT CARD PAYMENT

___ VISA ___ MASTERCARD ___ AMERICAN EXPRESS

CARD NUMBER_____

Expir. Date _____ Signature _____

PRINT CLEARLY – THIS IS YOUR MAILING LABEL

SHIP TO: Phone No. _____

Name _____

Address _____

_____

City _____ State _____ Zip _____

Expedited Service (enter service required): _____

Send the completed form and your check, made out to the **American Psychological Association**, or your credit card information to:

APA Order Department
750 First Street, NE
Washington, DC 20002-4242

All orders must be prepaid. Allow 4-6 weeks after the journal is published for delivery of a single copy or the first copy of a subscription.

Professional Psychology: Research and Practice
2009, Vol. ●●, No. ●, 000–000

0735-7028/09/$12.00 DOI: 10.1037/a0015808

An Effect Size Primer: A Guide for Clinicians and Researchers

Christopher J. Ferguson
Texas A&M International University

Increasing emphasis has been placed on the use of effect size reporting in the analysis of social science data. Nonetheless, the use of effect size reporting remains inconsistent, and interpretation of effect size estimates continues to be confused. Researchers are presented with numerous effect sizes estimate options, not all of which are appropriate for every research question. Clinicians also may have little guidance in the interpretation of effect sizes relevant for clinical practice. The current article provides a primer of effect size estimates for the social sciences. Common effect sizes estimates, their use, and interpretations are presented as a guide for researchers.

Keywords: effect size (statistical), null hypothesis testing, experimentation, statistical analysis, statistical significance, practical significance

By the 1990s, statisticians had been aware for some time that null-hypothesis significance testing (NHST) was, in many respects, insufficient for interpreting social science data (Berkson, 1938; Cohen, 1994; Loftus, 1996; Lykken, 1968; Meehl, 1978; Snyder & Lawson, 1993). Subsequently the Wilkinson Task Force (Wilkinson & Task Force on Statistical Inference, 1999) recommended the reporting of effect sizes and effect size confidence intervals (CIs). Nonetheless, the use of effect size measures remains inconsistent (Fidler et al., 2005; Osborne, 2008; Sink & Stroh, 2006). Researchers and clinicians may find themselves with little guidance as to how to select from among a multitude of available effect sizes, interpret data from research, or gauge the practical utility of reported effect sizes. The current article seeks to provide a primer for clinicians and researchers in understanding effect size reporting and interpretation.

The Purpose of Effect Size Reporting

NHST, has long been regarded as an imperfect tool for examining data (e.g., Cohen, 1994; Loftus, 1996). Statistical significance of NHST is the product of several factors: the "true" effect size in the population, the size of the sample used, and the alpha (p) level selected. Limitations of NHST include sensitivity to sample size, inability to accept the null hypothesis, and the failure of NHST to determine the practical significance of statistical relationships (Cohen, 1992, 1994; Loftus, 1996; Osborne, 2008). Kirk (1996) puts the limitations of NHST succinctly in noting that they fall under three main categories:

First, NHST does not adequately answer research questions. Regarding falsify-ability, scientists need to know the probability that a null hypothesis is true, given a data set. Unfortunately, NHST tells us the opposite, namely how likely a data set is to have occurred, given that the null hypothesis is true (Cohen, 1994; Kirk, 1996).

Second, no two sample means are ever identical (Tukey, 1991). The null hypothesis is, on a microscopic level at least, always false (Kirk, 1996). The result is the quixotic quest for power to demonstrate any difference as statistically significant without considering whether small differences are meaningful. This is particularly an issue when sample selection is nonrandom as sampling error is underestimated in NHST when sampling is nonrandom. NHST risks becoming something of a "trivial exercise" as a result (Kirk, 1996, p. 747).

Third, the .05 *p* level is arbitrary, leading researchers to come to different conclusions from equal treatment effects (Kirk, 1996). A researcher who finds that a treatment effect is nonsignificant using a sample of 100 participants, randomly assigned, may find that simply adding 100 more participants produces statistically significant effects, even though the treatment effects remain identical. This criticism is put most eloquently in Rosnow and Rosenthal's (1989) famous quote "Surely God loves the .06 nearly as much as the .05" (p. 1277).

At present, no clear replacement for NHST has emerged. However, the Wilkinson Task Force (1999) recommends the use of effect size in addition to NHST.

Effect sizes estimate the magnitude of effect or association between two or more variables (Ferguson, in press; Snyder & Lawson, 1993). As with all statistical tools, effect size estimates are just that, estimates. Mostly, effect sizes are resistant to sample size influence, and thus provide a truer measure of the magnitude of effect between variables.

Effect sizes seen in the social sciences are oftentimes very small (Rosnow & Rosenthal, 2003). This has led to difficulties in their interpretation. There is no agreement on what magnitude of effect is necessary to establish practical significance. Cohen (1992) of-

CHRISTOPHER J. FERGUSON received his PhD from the University of Central Florida. He is currently an assistant professor of clinical and forensic psychology at Texas A&M International University. His primary research interests focus on violent behavior, youth violence as well as positive and negative consequences of violent video game exposure. He is also interested in measurement issues in psychology, and ways in which hypotheses are tested and evaluated.

CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to Christopher J. Ferguson, Department of Behavioral, Applied Sciences and Criminal Justice, Texas A&M International University, 5201 University Boulevard, Laredo, TX 78041. E-mail: [illegible]Ferguson1111@aol.com

tapraid5/z2h-profes/z2h-profes/z2h00309/z2h2269d09z | xppws | S=1 | 3/30/09 | 4:38 | Art: 2009-0620

fers the value of $r = .1$, as a cut-off for "small" effects (which would indicate only a 1% overlap in variance between two variables). However, Cohen did not anchor his recommendations across effect sizes; as such, his recommendations for r and d ultimately differ in magnitude when translated from one to another. For instance, Cohen suggests that $r = .3$ and $d = .5$ each indicate a cut-off for moderate effects, yet $r = .3$ is not the equivalent of $d = .5$. Other scholars suggest a minimum of $r = .2$ (Franzblau, 1958; Lipsey, 1998) or .3 (Hinkle, Weirsma, & Jurs, 1988). In the current article, all effect size recommendations, where possible, are anchored to a minimum of $r = .2$, for practical significance (Franzblau, 1958; Lipsey, 1998). These readily convert from r to d for instance, without altering the interpretation. Note that this is a suggested minimum not a guarantee that observed effect sizes larger than $r = .2$ are practically significant. Such cut-offs are merely guidelines, and should not be applied rigidly (Cohen, 1992; Snyder & Lawson, 1993; Thompson, 2002). T1 Table 1 presents suggestions for effect size interpretation based on several previous reviews (Franzblau, 1958; Lipsey, 1998), although scholars are cautioned that effect size interpretation should be context specific, weighing the merits and potential unreliability of the study methodology against potential real-world impact of small or large effects.

Effect sizes can be thought of as falling into four general categories (Kline, 2004; Vacha-Haase & Thompson, 2004):

1. Group difference indices. As the name implies, these estimates usually note the magnitude of difference between two or more groups. Cohen's d is an example here.

2. Strength of association indices. These estimates usually examine the magnitude of shared variance between two or more variables. Pearson r is an example.

3. Corrected estimates. These measures, such as the adjusted R^2 correct for sampling error because of smaller sample sizes.

4. Risk estimates. These measures compare relative risk for a particular outcome between two or more groups. More commonly used in medical outcome research, these include relative risk (RR) and odds ratio (OR).

Group Difference Indices

Group difference indices lend themselves nicely to categorical or experimental outcomes rather than continuous or correlational data. The most commonly used such measure is Cohen's d (Cohen, 1969). Cohen's d is a rather simple statistical expression, namely the difference between two group outcomes divided by the population standard deviation. This is represented in the following formula: $d = (\mu 1 - \mu 2)/\sigma$.

The population standard deviation is an unknown, which leads to concerns how it should be represented. Often the treatment and control group standard deviations are pooled, although this results in problems in the event of multiple comparisons, such as in ANOVA designs. The population standard deviation estimate may vary from one contrast to another in such designs.

Glass proposed an alternate called delta (Δ), which substitutes the control group's standard deviation for the population standard deviation, thus standardizing the denominator variable across experimental contrasts: $\Delta = (\mu 1 - \mu 2)/SD_{\text{Control}}$.

Delta assumes an experimental design with a default control group that is representative of the population from which the samples are drawn. Delta would be less applicable for contrasts in which no clear control group exists (e.g., comparing males vs. females).

Hedges (1981) proposed a slightly different fix, in his statistic g, by pooling the standard deviations of all experimental groups and the control group resulting in a single standard deviation estimate. Although the resultant estimate is not really a great estimate of σ, it is more standardized than either d or Δ and as such may represent the best option for ANOVA designs.

Sapp, Obiakor, Gregas, and Scholze (2007) discuss a multivariate alternative for d with use with multiple dependent variables. This is represented by the formula $D^2 = (y_1\text{-}y_2)'S^{-1}(y_1\text{-}y_2)$, where $(y_1\text{-}y_2)$ is a vector of means and S^{-1} is the inverse of the sample covariance matrix. It is recommended that individuals interested in using D^2 consult Sapp, Obiakor, Gregas, and Scholze (2007) for a detailed discussion.

Strength of Association Indices

Strength of association indices estimate the shared variance between two or more variables. Arguably, they are more accurate for continuous data than are indices such as d, while maintaining usefulness is representing the effect size for categorical and experimental data. Most experimental outcomes using t or F statistics can be readily converted to the Pearson r (see Rosenthal & DiMatteo [2001] and Rosnow & Rosenthal [2003] for relevant formulas).

The most commonly used strength of association measure is Pearson's r that indicates the degree of shared variance between

Table 1
Effect Size Interpretation Suggestions for Social Science Data

Type of effect size estimate	Included indices	RMPE	Moderate effect	Strong effect
Group difference	d, Δ, g	.41	1.15	2.70
Strength of association	r, R, φ, ρ, partial r, β, r_h, tau	.2	.5	.8
Squared association indices	r^2, R^2, η^2, adjusted R^2, ω^2, ϵ^2	.04	.25	.64
Risk estimates	RR, OR	2.0*	3.0	4.0

Note. RMPE = recommended minimum effect size representing a "practically" significant effect for social science data. For effects with highly valid dependent measures (e.g., death) and using rigorous controlled outcome trials, lower values may have practical value. RR = relative risk; OR = odds ratio.
* These are not anchored to r and should be interpreted with caution.

two variables. R, R^2, η^2, φ, and Cramer's V are all related indices. R and R^2 in their unadjusted form are largely multiple regression equivalents of r and r^2 and generally indicate the strength of association between one dependent variable and multiple predictors. Cramer's V is typically used to represent the strength of association from chi-squared analyses as represented by the following formula: $V = [\chi^2/N * (k - 1)]^{1/2}$. In the event of a 2×2 table, the similar φ statistic is used: $\varphi = [\chi^2/N]^{1/2}$.

In this formula, k represents the minimal number of either rows or columns in the chi-squared display. One cautionary note about φ is in order. Some scholars have attempted to use φ to calculate effect size estimates from binomial medical epidemiology research (Rosnow & Rosenthal, 2003). This is a flawed approach as it corrupts the treatment effect size with that of the disease's own effect, producing a largely irrelevant interaction term effect size, not the effect size for treatment itself (Crow, 1991; Ferguson, in press; Hsu, 2004; Kraemer, 2005). Attempting to compute φ from such data results in a severely attenuated effect size estimate. This approach has been long recognized as flawed (Crow, 1991; Hsu, 2004; Kraemer, 2005), yet continues to be used.

As an example, Rosnow and Rosenthal (2003) discuss the effectiveness of the Salk vaccine on polio. From a treatment group (n = 200,745) receiving the Salk vaccine, 33 later had polio develop. From a control group (n = 201,229), polio developed in 115. Although the relative risk of not receiving the Salk vaccine is a moderate 3.48, the authors calculate φ = .011 (square root of χ^2 = 45.52 divided by n = 401,975), barely different from no effect at all. What if Salk was 100% effective in preventing polio (which should conceptually correspond to r = 1.00)? Observing the maximum φ, we observe that this barely changes with φ_{max} = .017 (χ^2 = 114.74; n = 401,975). In other words, the range of possible values for φ in the Salk vaccine trial instead of ranging from 0 to 1.00 as they should, ranges only from 0 to .017. This is clearly problematic. The problem is that the formula uses sample size. Dividing any value of χ^2 by n = 401,975 will result in artificially low scores. Because the epidemiologists who studied the Salk vaccine have no way of knowing who might be exposed to polio in advance, they used a wide "net" sampling approach. Most of the resultant sample were never exposed to polio, and thus are irrelevant to the hypothesis "How effective is the Salk vaccine in preventing polio in individuals exposed to polio". The r_h method (Ferguson, in press) accounts for this sampling problem by comparing only those individuals likely to have been exposed to polio. From the control group, we know that approximately 115 individuals can be expected to have polio develop. If Salk is completely ineffective, we would expect approximately 115 to have polio develop in the treatment groups as well (adjusting for any differences in sample size). However, only 33 have polio develop and 82 do not. Developing a binomial effect size display from these data, we find that for individuals actually likely to have been exposed to polio, r = .744 (χ^2 = 127.43; n = 230), not .011. This effect size is likely a better index of the actual effectiveness of the Salk vaccine and makes more sense given an RR of 3.48.

One further note about r is that the Pearson r works best with continuous and normally distributed data. It is important that the proper correlational method is used, otherwise resultant effect sizes will likely be truncated. For instance, ordinal data, or data that is nonnormally distributed, may be better handled by the Spearman rho (ρ) than Pearson r. Overall, it is argued here that r is a better effect size indicator for most psychological data than is d, as it provides the most accurate representation of continuous as well as categorical data, is well known and easy to interpret. As a cautionary note Wang and Thompson (2006) have indicated that r can be positively biased, particularly with small samples. They suggest using either the Ezekiel formula $1-[(n - 1)/(n - p - 1)][1 - R^2]$ or Smith formula $1-[n/(n - p)](1 - R^2)$ as corrections. The Wherry $1-[(n - 1)/(n - k - 1)] * (1 - R^2)$ and Lord formulas $1-(1 - R^2) * [(n + k + 1)/(n - k - 1)]$ (see Snyder & Lawson, 1993) can also be used to replace R^2 (in both formulas k represents the number of predictor variables.) The Lord formula is the more conservative of the two, and some scholars have cautioned against use of Wherry (Sapp, 2006). In place of Wherry, the Herzberg formula $1 -[(n - 1)/(n - k - 1)]*[(n + k + 1)/n](1 - R^2)$ may be supplemented (Sapp, 2006).

Pearson r and d can be readily converted from one to the other using the following formulas:

$$r = [d^2/(d^2 + 4)]^{1/2}$$

$$d = 2r/(1 - r^2)^{1/2}$$

Eta-squared (η^2) is a commonly reported index. η^2 is typically represented as a ratio of variance terms (Nunnally & Bernstein, 1994; Sink & Stroh, 2006):

$$SS_{between}/SS_{total\ Or:}\sigma^2_{true}/\sigma^2_{total}$$

η^2 (or partial η^2) is most commonly used for factorial ANOVA designs, owing to its standardized availability through SPSS. η^2 also is generally a better index of curvilinear relationships than is r, with the result that η^2 estimates tend to be slightly higher in magnitude than r^2 estimates. η^2 estimates are interpreted more or less similarly to r^2 estimates with the size of the relationship indicating shared variance between two variables, or explained by an interaction term. Levine and Hullett (2002) caution that η^2 estimates may be inflated and should not be taken as equivalent to r^2. Researchers are cautioned to note as well that SPSS printouts may misreport partial η^2 estimates as η^2 (Levine & Hullett, 2002). Partial η^2 estimates are nonadditive (meaning they can potentially sum to greater than 100% of total variance explained). Sapp (2006) indicates that η^2 involves division by total sample size. Partial η^2 involves division by sample size minus number of groups. As such with large samples, the distinction between η^2 and partial η^2 will tend to be small.

Corrected estimates include effect size estimates that attempt to correct either for error, or shared variance because of other predictors. The former group includes adjusted R^2, Hays' ω^2, and ε^2, although many of these effect size estimates see only seldom use (Kirk, 2006). The latter category includes partial r, and standardized regression coefficients (β), which estimate the shared variance between two variables once variance because of other variables is controlled. Levine, Weber, Park, and Hullett (2008) express wariness about using β as the effect sizes seen tend to be lower than r. Similarly, partial r presents the strength of association once variance because of a third variable has been removed. For instance if one wishes to study the relationship between video game playing and aggression, it may be necessary to remove variance because of sex, as males both play more video games and are more aggressive (Ferguson, 2007). The resultant partial r

provides a better index of the unique variance shared by video games and aggression, once male sex is removed from the equations. In that sense, β and partial *r* may be more accurate than bivariate *r* as an effect size estimate by eliminating extraneous variance from a relationship that may artificially inflate nonadditive effect size estimates.

ω^2, and ε^2 are most commonly used for ANOVA designs, and thus can replace *d* or η^2. Indeed, they are recommended as a replacement for η^2 in particular (Levine & Hullett, 2002).

$$\omega^2 = [SS_{effect} - df_{effect} \times MSE]/[SS_{total} + MSE]$$

$$\varepsilon^2 = [SS_{effect} - (df_{effect} - 1)(MSE)]/SS_{total}$$

ε^2 tends to be larger than ω^2. Both are smaller in magnitude than r^2 or η^2. They are, mostly, more conservative estimates of effect size and, as such, highly recommended as replacements, particularly for η^2. Note that the ω^2 formula for random effects models is different from the fixed effects model presented here and can be found on page 340 of Snyder and Lawson (1993).

Several indices of effect size can be used with rank-order data. The Spearman rho (ρ) mentioned earlier is one of these. Kendall's τ provides an effect size estimate based on the number of concordant pairs in two rank orders (Kendall, 1938). Gamma is a somewhat similar measure, measuring the strength of association between concordant rank ordered pairs (Sapp, 2006). Somer's d is similar to Gamma, although asymmetrical in nature with presumed cause and effect variables. All four measures have value ranges identical to *r* although they are superior for rank-order data.

Risk Estimates

Risk estimate measures are more commonly used in medical research. They estimate the difference in risk for a particular outcome between two or more groups of individuals. Three main risk estimates in use include RR, OR, and risk difference (RD). Table 2 refers to the binomial display used in biomedical treatment outcome research.

All risk estimates are formed from ratios of individuals in each of the cells presented in Table 2. RR is a ratio of patients in a control group who contract an illness or condition to those in the treatment group who contract the same condition and is represented as (Bewick, Cheek, & Ball, 2004; Rosnow & Rosenthal, 2003):

$$RR = [A/(A + B)]/[C/(C + D)]$$

An RR of 1.0 indicates no difference in risk between the two groups. Below 1.0 indicates less risk for the control group than treatment group ("control" group is used broadly, indicating any group assumed in the hypothesis to have the greater risk and should not be taken to restrict RR to experimental designs only). An RR of 2.0 would indicate that the "control" group is twice as likely to demonstrate a given condition than the "treatment" group (thus, they have 100% greater risk). As an example, the American Cancer Society (2008) estimates the RR for smoking and lung cancer to be approximately 23 for male smokers. This means that male smokers are 23 times more likely to have lung cancer develop than male nonsmokers.

Table 2
Binomial Effect Size Display for Risk Estimates

Variable	Patients sick	Patients well
Control	A	B
Treatment	C	D

Note. The capitol letters represent the number or frequency of patients in each cell.

Just as with effect sizes seen more commonly in the social sciences, interpretation of RR can be context specific, depending upon the initial base risk of the disease. However, RR values between 1.0 and 2.0 are unlikely to have much practical value, and these small effects are likely to be highly unreliable and contradictory across studies. Indeed a plethora of observational studies with small RR effects have linked *a little of this with a little of that* in terms of purported harmful (or beneficial) factors on health that ultimately were found to be of little value (Smith & Ebrahim, 2001). However, researchers are reminded that RR must be interpreted by examining the actual base risk of the treatment group (i.e., lower risk group). An RR of 2.0 might not be terribly meaningful if the base risk is 1% but may be more meaningful if the base risk is 10%.

RR is likely a better estimate of effects for binomial data than are efforts to translate the same data into *r*. As such, it is recommended that psychological researchers become more familiar and comfortable with the use of RR as it may be superior to *r* for certain kinds of data, particularly binomial data.

The OR is a similar effect index using the following formula (Bewick, Cheek, & Ball, 2004; Rosnow & Rosenthal, 2003):

$$OR = (A/B)/(C/D)$$

OR represents the odds of control patients contracting a condition in proportion with treatment patients contracting a condition. So long as the risks of the disease are low, OR will approximate RR (Bewick, Cheek, & Ball, 2004). Davies, Crombie, and Tavakoli (1999) note that this approximate relationship breaks down when the risk in either group rises above 20%, with OR and RR becoming increasingly disparate. The problem with the OR as indicated by Rosnow and Rosenthal (2003) is that when the denominator terms become very small (i.e., few people survive or are unaffected), the OR can potentially accelerate to infinity. ORs greater than 3.0 (or less than 0.33 when the control group is less affected) are considered to be moderately strong (Haddock, Rindskopf, & Shadish, 1998). As with RR, the base risk must be considered with interpreting OR.

RD represents the difference between the proportion of control patients who contract a condition and the proportion of treatment patients who contract a condition. It is represented by the following formula (Bewick, Cheek, & Ball, 2004; Rosnow & Rosenthal, 2003):

$$RD = A/(A + B) - C/(C + D)$$

RD is fairly easy to interpret, representing the actual difference in risks between two groups. For instance RD = .04 represents a 4% risk difference between the control and treatment groups. Evaluating the clinical or practical significance of *RD* is trickier than with other measures. For instance, RD = .04 may be clini-

cally insignificant if the compared risks are 76% and 80%, but very significant if the compared risks are 1% and 5%.

Other Indices

Counternull

The counternull value of the effect size is the nonnull magnitude of effect that is equally supported by the data as a null effect itself (Rosenthal & Rubin, 1994). For most group difference effect sizes, this value is simply twice the observed effect size. So, for $d = .2$ with a nonsignificant result, the actual effect in the population is as equally likely to be $d = .4$ as it is $d = .0$. The counternull is intended to remind researchers that a nonsignificant result does not mean zero effect size, and that significant results do not necessarily mean practical importance.

Number Needed To Treat

The number needed to treat (NNT; Pinson & Gray, 2003) refers to the number of individuals who would need to receive treatment to produce one more positive response. For instance, a NNT of three would indicate that three patients would need to receive treatment for one more patient to show improvement than would have under no treatment. NNT works best with dichotomous outcomes (i.e., success/failure). NNT is influenced by the baseline rate of the disease in the population of interest (Pinson & Gray, 2003). NNT, as such provides a reasonably intuitive indices for evaluating the outcome of clinical trials. For example, if 70% of patients did not respond to a placebo treatment and 20% did not respond to an experimental treatment, this difference would be 50% or .5. NNT is the inverse of this $1/.5 = 2$, meaning that two patients would need to receive treatment for one more positive response than under placebo.

A Note on Practical Significance

In many respects, effect size estimates provide greater information for judgments about practical significance than does NHST. Yet, judgments about practical significance need to consider several issues. The quality of measurement and sampling strategies both have the potential to inflate or attenuate effect size estimates. For instance, an r^2 of .01 is likely to have much greater practical impact when the outcome is death or serious morbidity. The same effect size for a weakly validated interval/ratio-scale measure of behavior within individuals may be essentially nonnoticeable in the "real world." In addition, researchers may need to consider the risks and benefits of a particular intervention. An intervention with a weak effect size but no risks may be valuable. That same intervention may be less desirable if the risks are considerable. An intervention that causes a 1% decline in the rate of death for a particular disease may be valuable no matter what the risks (because no one can be deader than dead). By contrast, a depression treatment that reduces depressed mood only 1% may not be worth the financial costs accrued to patients.

A recent clinical trial of circumcision used to prevent HIV among Kenyan men found that of 1,391 men randomized to circumcision, 22 later had HIV, whereas 47 in the control uncircumcised group of 1,393 became HIV positive (Bailey et al., 2007). This is approximately a 53% reduction in risk, corresponding to an RR for uncircumcised males of 2.13 (95% CI 1.30-3.53). Although the effect size is small, the authors considered these results significant enough to stop the trial and offer circumcision to men in the control group. As such, it is important to use some critical thinking and perspective when evaluating practical significance. Bailey et al. (2007) represents a well-run randomized clinical outcome trial with a highly valid dependent variable (HIV-positive serum test). The interpretation applied to Bailey et al. should not be assumed to apply to all psychological studies, particularly with measures for which validity is weak or not known, or clinical cut-offs with established sensitivity and specificity are absent.

Some General Recommendations

The following guidelines are offered to assist in the choice of effect size measures and their interpretation.

1. Guidelines for the interpretation of many effect sizes are offered as part of this article. As noted by numerous scholars (e.g., Cohen, 1992; Snyder & Lawson, 1993; Thompson, 2002), rigid adherence to arbitrary guidelines is not recommended. However, this admonition should not be taken as license for the overinterpretation of weak effect sizes. Guidelines are suggested as minimum cut-offs, not guarantees that effect sizes exceeding those cut-offs are meaningful. Study limitations, failure to control for other relevant predictors or threats to internal validity, the reliability and validity of responses to the measures used, etc., should be considered when interpreting effect size.

2. Corrected effect sizes are preferable to uncorrected effect sizes. With the exception of adjusted R^2, these unfortunately require more by-hand calculation on the part of experimenters. Nonetheless, the corrected effect size estimates presented are likely to be more accurate estimates of the effect size in the population of interest.

3. For correlational designs, partial r and standardized regression coefficients are superior to bivariate r as they estimate the unique variance because of a predictor controlling for other variables. These estimates help to reduce nonadditive illusory correlations (that can sum to greater than 100% of explained variance) that overestimate the unique predictive value of variables.

4. Efforts to translate RR or OR into r or d should be discontinued. If such translations absolutely must be done, only r_h should be used as an effect size estimate when the data are binomial.

5. ρ, Somer's d or Kendall's τ should be used instead of r or d for ordinal data.

6. It is recommended that effect size CIss be reported along with effect size estimates (Sapp, 2004; Wilkinson et al., 1999). Effect size CIs that cross zero suggest that the null hypothesis should not be rejected (Sapp, 2004). Thus, the use of CIs provides more information than point estimates alone, and interpretations of effect size estimates should include analysis of their CI.

7. Although effect size is meant to represent a "true" effect in the population, it is important to understand that effect size estimates can be influenced by sampling and measurement. Samples that are too small, or that are nonrandom may produce biased effect size estimates that should be interpreted with caution. Various methodological issues can also influence effect size. Sample responses that demonstrate poor reliability on a measure may lower effect size estimates. By contrast, poorly standardized measures may inadvertently result in inflated effect size estimates if they allow researchers to select from multiple potential outcomes (Ferguson, 2007). Effect sizes also can vary depending upon the statistical methods used. Statistical methods that reduce variance, such as using false dichotomies, may truncate effect size estimates. Research designs that fail to control for extraneous variables will tend to produce higher effect sizes than those that adequately control extraneous variables if those variables are highly correlated with the independent and dependent measures (Olejnic & Algina, 2003). The use of parametric statistics when the assumptions of such tests have been violated may also produce biased effect size estimates. For instance, even with medical epidemiological studies, it has been observed that correlation outcomes, particularly those with weak effect sizes, often prove unreliable under scrutiny (Smith & Ebrahim, 2001).

Conclusion

This article presents a number of commonly used and important effect size measures. It is worth noting that it is not possible to describe all effect size measures available as they are quite numerous. There are several resources to which the reader is referred for further discussion (e.g., Cohen 1992; Sapp, 2006).

The current article is intended as an introductory primer for researchers who may not be familiar with the range of effect size estimates available to them, and when they should be used. No effect size measure is perfect for all situations, and some disagreement persists over the advantages of some over others. This article is intended as a general guideline. Ultimately, researchers are encouraged to select wisely (and conservatively) from among the options presented. Interpretation of effect sizes will always remain context specific. Accepting effect sizes of any magnitude as clinically or practically significant renders their usefulness as moot. Psychologists must be willing to objectively identify minute effects and interpret them as such. Only by being realistic will we know when we are actually on to something.

References

American Cancer Society. (2008). *Smoking and cancer mortality table.* Retrieved November 2, 2008, from http://www.cancer.org/docroot/PED/content/PED_10_2X_Smoking_and_Cancer_Mortality_Table.asp

Bailey, R., Moses, S., Parker, C., Agot, K., Maclean, I., Krieger, J., et al. (2007). Male circumcision for HIV prevention in young men in Kisumu, Kenya: A randomised controlled trial. *Lancet, 369,* 643–656.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association, 33,* 526–542.

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 11: Assessing risk. *Critical Care, 8,* 287–291. Retrieved January 11, 2008, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=15312212

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* 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.

Crow, E. (1991). Response to Rosenthal's comment, "How are we doing in soft psychology?" *American Psychologist, 46,* 1083.

Davies, H., Crombie, I., & Tavakoli, M. (1999). When can odds ratios mislead? *British Medical Journal, 316,* 989–991.

Ferguson, C. J. (2007). Evidence for publication bias in video game violence effects literature: A meta-analytic review. *Aggression and Violent Behavior, 12,* 470–482.

Ferguson, C. J. (in press). Is psychological research really as good as medical research? Effect size comparisons between psychology and medicine. *Review of General Psychology.* AQ: 1

Fidler, F., Cumming, G., Thomason, N., Pannuzzo, D., Smith, J., Fyffe, P., et al. (2005). Toward improved statistical reporting in the Journal of Consulting and Clinical Psychology. *Journal of Consulting and Clinical Psychology, 73,* 136–143.

Franzblau, A. (1958). *A primer of statistics for non-statisticians.* New York: Harcourt, Brace & World.

Haddock, C., Rindskopf, D., & Shadish, W. (1998). Using odds ratios as effect sizes for meta analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3,* 339–353.

Hedges, L. (1981). Distributional theory for Glass' estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128.

Hinkle, D., Weirsma, W., & Jurs, S. (1988). *Applied statistics for the behavioral sciences.* Boston: Houghton Mifflin.

Hsu, L. M. (2004). Biases of success rate differences shown in binomial effect size displays. *Psychological Bulletin, 9,* 183–197.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika, 30,* 81–89.

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

Kline, R. B. (2004). *Beyond significance testing.* Washington, DC: American Psychological Association.

Kraemer, H. C. (2005). A simple effect size indicator for two-group comparisons?: A comment on $r_{equivalent}$. *Psychological Methods, 10,* 413–419.

Levine, T., & Hullett, C. (2002). Eta squared, partial eta squared and misreporting of effect size in communication research. *Human Communication Research, 28,* 612–625.

Levine, T., Weber, R., Park, H., & Hullett, C. (2008), A communication researcher's guide to null hypothesis significance testing and alternatives. *Human Communication Research, 34,* 188–209.

Lipsey, M. W. (1998). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 39–68). Thousand Oaks, CA: Sage.

Loftus, G. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science, 5,* 161–171.

Lykken, D. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70,* 151–159.

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

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory.* New York: McGraw-Hill.

Olejnic, S., & Algina, J. (2003). Generalized eta and omega squared

statistics: Measures of effect size for some common research designs. *Psychological Methods, 8,* 434–447.

Osborne, J. (2008). Sweating the small stuff in educational psychology: How effect size and power *reporting* failed to change from 1969 to 1999, and what that means for the future of changing practices. *Educational Psychology, 28,* 151–160.

Pinson, L., & Gray, G. (2003). Number needed to treat: An underused measure of treatment effect. *Psychiatric Services, 54,* 145.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45,* 775–777.

Rosenthal, R., & DiMatteo, M. (2001). Meta analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52,* 59–82.

Rosenthal, R., & Rubin, D. (1994). The counternull value of an effect size: A new statistic. *Psychological Science, 5,* 329–334.

Rosnow, R., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276–1284.

Rosnow, R., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology, 57,* 221–237.

Sapp, M. (2004). Confidence intervals within hypnosis research. *Sleep and Hypnosis, 6,* 169–176.

Sapp, M. (2006). *Basic psychological measurement, research designs, and statistics without math.* Springfield, IL: Charles C. Thomas Publisher.

Sapp, M., Obiakor, F., Gregas, A., & Scholze, S. (2007). Mahalanobis distance: A multivariate measure of effect in hypnosis research. *Sleep and Hypnosis, 9,* 67–70.

Sink, C., & Stroh, H. (2006). Practical significance: The use of effect sizes in school counseling research. *Professional School Counseling, 9,* 401–411.

Smith, G., & Ebrahim, S. (2001). Epidemiology: Is it time to call it a day? *International Journal of Epidemiology, 30,* 1–11.

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

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.

Tukey, J. (1991). The philosophy of multiple comparisons. *Statistical Science, 6,* 100–116.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling and Psychology, 51,* 473–481.

Wang, Z., & Thompson, B. (2007). Is the Pearson r^2 biased, and if so, what is the best correction formula? *Journal of Experimental Education, 75,* 109–125.

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

Received January 5, 2009
Revision received February 24, 2009
Accepted March 2, 2009 ■

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES 1

AQ1: Author: Please update the following references, if available: Ferguson (in press).

AQ2: Author: Please cite Rosenthal (1990) in the text or remove it from the list.