AMERICAN COUNSELING ASSOCIATION

Professional Counseling Digest

ACAPCD-26 2009

A Primer in Evaluating Quantitative Research for Counseling Professionals

Richard S. Balkin, Ph.D. and Donna S. Sheperis, Ph.D.

Counselors often report increased stress and low self-efficacy in understanding and interpreting statistical research (Pan & Tang, 2004). Many counselors avoid or merely scan statistical research because of a lack of familiarity with quantitative designs. However, it is important to recognize that counselors have an ethical responsibility to develop an understanding of research. In fact, ACA's Code of Ethics (2005) specifically directs counselors to "engage in counseling practices that are based on rigorous research methodologies" (p. 9). This entry serves as a basic overview of quantitative research that will aid in both reducing anxiety and increasing counselors' understanding of applications of quantitative research.

Generalizability: Identifying Who Is Helped

Most counseling journals require reporting demographic characteristics (e.g., gender, race, and ethnicity) of participants. Demographic information clarifies the various attributes of the participants and gives the reader an understanding of *generalizability*—how the study translates across a population. However, population validity, internal experimental validity, and/or external experimental validity are essential elements to consider before assuming generalizability. *Population validity* refers to the manner in which participants were selected and represent the population (Gall, Gall, & Borg, 2003). Ideally, counseling researchers should use random sampling. In practice, most counseling researchers rely on convenience sampling (e.g., many counseling studies are conducted with a college student population and then generalized to a larger adult population). What is pertinent to most practicing counselors is whether the participants in the study are representative of the target population (e.g., clients; Balkin, 2010). Realistically, the study is only meaningful if the participants are generalizable to a larger population.

For experimental studies, *internal experimental validity* is the extent to which changes in the dependent variable are due to the treatment or intervention rather than other factors (Balkin, 2010). *External experimental validity* is the extent to which the design of the research translates to viable conclusions about the participants outside of the experimental setting (Balkin, 2010). For example, a client in a psychiatric hospital may learn and implement new coping skills in the experimental setting of the hospital. Upon discharge, that same client may have difficulty implementing the new coping skills, because the structure and environment in the hospital do not necessarily translate to an external setting.

The Statistical Test—What Does It Really Mean?

The interpretation of statistical tests is straight forward. Yet this is often where counselors get overwhelmed or become misguided in reading research articles. Many believe that a result that is statistically significant is, by default, important and meaningful. The truth is that a statistical test is merely a measure of probability, not meaningfulness. Essentially, a statistically significant result tells us that achieving a similar relationship or difference is likely to occur again because it is not merely due to chance. Two aspects identify this probability: the alpha level and the *p*-value. The alpha level is an indication of how much *type-I error* researchers have allowed in the study. Type-I error is the likelihood of identifying relationships or differences when no relationship or difference truly exists. The *p*-value is an indication of the precise amount of type-I error in the study (e.g., if a researcher indicates an alpha level of .05, then there is a 5% chance that the researcher could make a type-I error). This level is set prior to the data analysis and is based on the researcher's willingness to accept a certain probability for error.

To illustrate this concept, consider a study where the *p* value is less than the alpha of 5%, (reported as $p < .05$). This *p*-value indicates that the chance of making a type-I error with the given result is less than 5%. Such a result is considered *statistically significant* because the found value (*p*) is less than the previously determined value (alpha). Often, researchers will provide an exact value for *p*, such as $p = .035$; this indicated a 3.5% chance of making a type-I error—still less than the stated 5%. Common statistical tests produce *p* values. Table 1 provides an overview of those common statistical tests in the social sciences.

The reporting of statistical results follow this format: a Roman or Greek letter, followed by a number or pair of numbers in parenthesis, followed by a value, and concluded by a comparison to *p* [e.g., $F(3, 96) = 13.81$, $p < .001$]. While the presented numbers may look complex, they are really a summary statement of the research results. The Roman or Greek letter represents the type of test. In this case an *F* indicates that an ANOVA was performed. The number(s) in parenthesis identifies *degrees of freedom* (i.e., an estimate of parameters or variability within a data set). The value after the

Table 1. Synopsis of Common Statistical Tests in Counseling Research

Type of Test	Purpose
z-test	tests hypotheses between a sample mean and a population mean
t-test	tests hypotheses between two sample means only
F-test	tests hypotheses between or among two or more sets of variables
χ^2	tests hypotheses between two or more sets of variables when utilizing nonparametric tests

equal sign is based on a calculation that incorporates changes in the dependent variable and error in measurement. The *p*-value indicates whether or not the result is statistically significant.

The results section of a study should address both statistical and practical significance. A common concern in research is the use of multiple *p* levels to report significance (e.g., $*p < .05$; $**p < .01$). Rules for hypothesis testing require that researchers select one alpha level, not several (Balkin & Erford, 2008).

Evaluating Meaningfulness: Effect Size

Because statistical significance is merely a term of probability, many counseling journals (e.g., *Journal of Counseling & Development*, *Counselor Education and Supervision*) have mandated the use of effect size to measure practical significance. *Effect size* is critical because it tells us the magnitude of the difference between or among variables. In general, effect size is reported in two methods: (a) variance accounted for in the model, and (b) differences based on standard deviation units (Trusty, Thompson, & Petrocelli, 2004). Common effect size measures for variance accounted for in the model include η^2,ω^2, R^2, λ, and various correlation coefficients (e.g., *r*, *rs*, *sr*, ϕ). Each of these effect size measures help researchers and consumers to determine the degree to which changes in the criterion or dependent variable(s) occur with respect to changes or manipulation in the predictor or independent variable(s). Differences also are reported based on standard deviation units, using such tests as Cohen's *d* or Cohen's *f*, which indicate the extent to which differences exist between or among groups.

There are many types of effect size. Thus the means of identifying and reporting the extent of the relationships or differences will vary depending on the study. Furthermore, not all journals mandate the reporting of effect size, leaving the reader without valuable information related to the magnitude of the differences or relationships found. General guidelines do exist in interpreting effect size measures, and researchers often report an interpretation of the effect size found. Cohen (1992) provided general interpretations of effect size classifying the measures in to *small*, *medium*, or *large*. Thus, the importance of the findings can be ascertained through these categories.

Effect size should be provided in the results section. An interpretative statement, often using Cohen's (1992) standards of *small, medium, and large* effect sizes, related to the meaningfulness of the finding(s) should be provided. "The manuscript is incomplete unless these effects are evaluated in the context of the study and in the larger context of knowledge" (Trusty et al., 2004, p. 109). While effect size may be unfamiliar to many counselors, it is easy to see why it is a critical addition to the results. Again, it signifies the difference between what is statistically significant and what is practically significant or meaningful.

As counseling journals continue to update standards related to publishing research, both those who create and those who utilize counseling research need to understand the role of sample selection, statistical significance, and practical significance. Before implementing research findings into clinical practice, you must give serious consideration to sampling and meaningfulness of the findings. Information on sampling should be in the methods section related to participants. Before deciding to implement findings into practice, similarities should be evident between the participants of the study and the clients within your practice.

In addition to similarity between a sample and client pool, evaluation of appropriate statistical tests and practical significance is important. Moreover, a reflection of the interpretative statements regarding effect size should be reviewed in the discussion section of a manuscript. Discussion should not be accepted at face value, because a common error in social science research is to ignore the effect size and emphasize the importance of the results solely based on statistical significance. Clearly understanding and implementing research findings responsibly are central to complying with ethical standards for providing the most effective treatment. Thus, a solid foundation to evaluate research is integral to upholding the highest standard of professional and ethical functioning as counselors.

References

American Counseling Association. (2005). *ACA Code of Ethics*. Alexandria, VA: ACA.

Balkin, R.S. (2010). Experimental designs. In C.J. Sheperis, J.S. Young, and H.M. Daniels (Eds.), *Counseling research: Quantitative, qualitative, and single-subject design*. Columbus, OH: Prentice Hall.

Balkin, R. S., & Erford, B. T. (2008). Statistical hypothesis testing. In B.T. Erford (Ed.), *Research and evaluation in counseling*. Boston: Lahaska Press.

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

Gall, M.D., Gall, J.P., & Borg, W.R. (2003). *Educational research: An introduction* (7th ed.). Boston, MA: Allyn & Bacon.

Pan, W., & Tang, M. (2004). Examining the effectiveness of innovative instructional methods on reducing statistics anxiety for graduate students in the social sciences. *Journal of Instructional Psychology, 31*, 149-159.

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

Richard S. Balkin, Ph.D., LPC-S, NCC, is an associate professor at Texas A&M University – Commerce. Donna S. Sheperis, Ph.D., LPC, is an assistant professor at Delta State University.

The authors would like to acknowledge Dr. Carl Sheperis for his assistance with this manuscript.

ACA Professional Counseling Digests are produced by Counseling Outfitters, LLC, in collaboration with the American Counseling Association.

Copyright © 2009 American Counseling Association.

Suggested APA style reference:
Balkin, R. S., & Sheperis, D. S. (2009). *A primer in evaluating quantitative research for counseling* (ACAPCD-26). Alexandria, VA: American Counseling Association.